

日 本 国 特 許 庁  
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出 願 年 月 日

Date of Application:

2002年 6月26日

出 願 番 号

Application Number:

特願2002-186690

[ ST.10/C ]:

[ JP2002-186690 ]

出 願 人

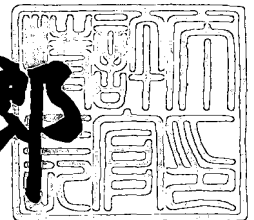
Applicant(s):

ソニー株式会社

2003年 5月23日

特 許 庁 長 官  
Commissioner,  
Japan Patent Office

太田信一郎



出証番号 出証特2003-3037902

【書類名】 特許願

【整理番号】 0290049715

【提出日】 平成14年 6月26日

【あて先】 特許庁長官殿

【国際特許分類】 G06T 1/00  
G10L 11/00

【発明者】

【住所又は居所】 東京都品川区北品川6丁目7番35号 ソニー株式会社  
内

【氏名】 近藤 哲二郎

【発明者】

【住所又は居所】 東京都品川区北品川6丁目7番35号 ソニー株式会社  
内

【氏名】 奥村 裕二

【発明者】

【住所又は居所】 東京都品川区北品川6丁目7番35号 ソニー株式会社  
内

【氏名】 藤島 幸一

【発明者】

【住所又は居所】 東京都品川区北品川6丁目7番35号 ソニー株式会社  
内

【氏名】 大月 知之

【特許出願人】

【識別番号】 000002185

【氏名又は名称】 ソニー株式会社

【代理人】

【識別番号】 100090376

【弁理士】

【氏名又は名称】 山口 邦夫

【電話番号】 03-3291-6251

【選任した代理人】

【識別番号】 100095496

【弁理士】

【氏名又は名称】 佐々木 榮二

【電話番号】 03-3291-6251

【手数料の表示】

【予納台帳番号】 007548

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9709004

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 観客状況推定装置と観客状況推定方法および観客状況推定プログラム

【特許請求の範囲】

【請求項 1】 観客を撮影した画像信号を生成する撮像手段と、  
前記画像信号に基づいて、前記観客の動き量を検出する動き量検出手段と、  
前記動き量に基づいて観客状況を推定する推定手段とを有する  
ことを特徴とする観客状況推定装置。

【請求項 2】 前記動き量検出手段は、前記画像信号に基づき画像の動きベクトルを求め、該動きベクトルの大きさの平均を示す平均動き量を、前記観客の動き量とする  
ことを特徴とする請求項 1 記載の観客状況推定装置。

【請求項 3】 前記動き量検出手段は、前記画像信号に基づき画像の動きベクトルを求め、該動きベクトルの大きさの平均を示す平均動き量を算出して、該平均動き量の時間方向の平均値である時間マクロ動き量を、前記観客の動き量とする  
ことを特徴とする請求項 1 記載の観客状況推定装置。

【請求項 4】 前記推定手段では、前記動き量が所定レベルよりも大きいとき、前記観客状況を手拍子状態あるいは手拍子状態と推定する  
ことを特徴とする請求項 1 記載の観客状況推定装置。

【請求項 5】 観客を撮影した画像信号を生成する撮像手段と、  
前記画像信号に基づいて、前記観客の動き周期性を検出する動き周期性検出手段と、  
前記観客の動き周期性に基づいて観客状況を推定する推定手段とを有する  
ことを特徴とする観客状況推定装置。

【請求項 6】 前記動き周期性検出手段は、前記画像信号に基づき画像の動きベクトルを求め、該動きベクトルの大きさの平均を示す平均動き量を算出して、該平均動き量の自己相関最大位置を検出し、該自己相関最大位置の分散を、前記動き周期性とする

ことを特徴とする請求項 5 記載の観客状況推定装置。

【請求項 7】 前記動き周期性検出手段は、前記画像信号に基づき画像の動きベクトルを求め、該動きベクトルの大きさの平均を示す平均動き量を算出して、該平均動き量における低周波周波成分比を前記動き周期性とする

ことを特徴とする請求項 5 記載の観客状況推定装置。

【請求項 8】 前記推定手段では、前記動き周期性が所定レベルよりも大きいときには前記観客状況を手拍子状態、前記所定レベルよりも大きくないときには前記観客状況を拍手状態と推定する

ことを特徴とする請求項 5 記載の観客状況推定装置。

【請求項 9】 観客からの音声に応じた音声信号を生成する音声取得手段と、

前記音声信号に基づいて、前記観客からの音声の音量を検出する音量検出手段と、

前記音量に基づいて観客状況を推定する推定手段とを有する  
ことを特徴とする観客状況推定装置。

【請求項 10】 前記音量検出手段は、前記音声信号の音声レベルの分散を前記音量とする

ことを特徴とする請求項 9 記載の観客状況推定装置。

【請求項 11】 前記音量検出手段は、前記音声信号の音声レベルの分散を算出して、該分散の時間方向の平均値である時間マクロ音量を前記音量とする  
ことを特徴とする請求項 9 記載の観客状況推定装置。

【請求項 12】 前記推定手段では、前記音量が所定レベルよりも大きいときには前記観客状況を笑い状態と推定する

ことを特徴とする請求項 9 記載の観客状況推定装置。

【請求項 13】 観客からの音声に応じた音声信号を生成する音声取得手段と、

前記音声信号に基づいて、前記観客からの音声の周期性を検出する音声周期性検出手段と、

前記音声の周期性に基づいて観客状況を推定する推定手段とを有する

ことを特徴とする観客状況推定装置。

【請求項 1 4】 前記音声周期性検出手段は、前記音声信号から自己相関最大位置を順次検出し、該自己相関最大位置の分散を、前記音声の周期性とすることを特徴とする請求項 1 3 記載の観客状況推定装置。

【請求項 1 5】 音量検出手段を設け、  
前記音声周期性検出手段は、前記音声信号に基づいて前記音量検出手段で検出された前記観客からの音声の音量における低周波成分比を、前記音声の周期性とすることを特徴とする請求項 1 3 記載の観客状況推定装置。

【請求項 1 6】 前記推定手段では、前記音声の周期性が所定レベルよりも大きいときには前記観客状況を拍手状態と推定することを特徴とする請求項 1 3 記載の観客状況推定装置。

【請求項 1 7】 観客からの音声に応じた音声信号を生成する音声取得手段と、  
前記音声信号のフーリエ変換を行い、該フーリエ変換結果の重心を検出するスペクトル検出手段と、  
前記重心に基づいて観客状況を推定する推定手段とを有することを特徴とする観客状況推定装置。

【請求項 1 8】 前記推定手段は、前記重心が所定レベルよりも大きいときには前記観客状況を手拍子状態あるいは拍手状態と推定することを特徴とする請求項 1 7 記載の観客状況推定装置。

【請求項 1 9】 観客からの音声に応じた音声信号を生成する音声取得手段と、  
前記音声信号に基づく前記のフーリエ変換を行い、該フーリエ変換結果に基づいて高周波対低周波成分比を算出する成分比検出手段と、  
前記高周波対低周波成分比に基づいて観客状況を推定する推定手段とを有することを特徴とする観客状況推定装置。

【請求項 2 0】 前記推定手段は、前記高周波対低周波成分比によって、前記低周波成分に対する前記高周波成分の割合が所定レベルよりも大きいときには

前記観客状況を手拍子状態あるいは拍手状態と推定することを特徴とする請求項 1 9 記載の観客状況推定装置。

【請求項 2 1】 観客を撮影した画像信号および／または前記観客からの音声に応じた音声信号を生成する入力手段と、

前記画像信号に基づき、前記観客の動き量や動き周期性をいずれか 1 つあるいは両方を検出し、前記音声信号に基づき、前記観客のからの音声の音量や前記音声の周期性や前記音声の周波数成分に関する情報のいずれか 1 つあるいは複数を検出する特徴量検出手段と、

前記特徴量検出手段での検出結果に基づいて、観客状況を推定する推定手段とを有することを特徴とする観客状況推定装置。

【請求項 2 2】 観客を撮影して画像信号を生成し、前記画像信号に基づいて、前記観客の動き量を検出し、前記動き量に基づいて観客状況を推定することを特徴とする観客状況推定方法。

【請求項 2 3】 前記画像信号に基づき画像の動きベクトルを求め、該動きベクトルの大きさの平均を示す平均動き量を、前記観客の動き量とすることを特徴とする請求項 2 2 記載の観客状況推定方法。

【請求項 2 4】 前記画像信号に基づき画像の動きベクトルを求め、該動きベクトルの大きさの平均を示す平均動き量を算出して、該平均動き量の時間方向の平均値である時間マクロ動き量を、前記観客の動き量とすることを特徴とする請求項 2 2 記載の観客状況推定方法。

【請求項 2 5】 前記動き量が所定レベルよりも大きいとき、前記観客状況を手拍子状態あるいは拍手状態と推定することを特徴とする請求項 2 2 記載の観客状況推定方法。

【請求項 2 6】 観客を撮影して画像信号を生成し、前記画像信号に基づいて、前記観客の動き周期性を検出し、前記観客の動き周期性に基づいて観客状況を推定することを特徴とする観客状況推定方法。

【請求項 2 7】 前記画像信号に基づき画像の動きベクトルを求め、該動きベクトルの大きさの平均を示す平均動き量を算出して、該平均動き量の自己相関最大位置を検出し、該自己相関最大位置の分散を、前記動き周期性とすることを特徴とする請求項 2 6 記載の観客状況推定方法。

【請求項 2 8】 前記画像信号に基づき画像の動きベクトルを求め、該動きベクトルの大きさの平均を示す平均動き量を算出して、該平均動き量における低周波成分比を前記動き周期性とすることを特徴とする請求項 2 6 記載の観客状況推定方法。

【請求項 2 9】 前記動き周期性が所定レベルよりも大きいときには前記観客状況を手拍子状態、前記所定レベルよりも大きくないときには前記観客状況を拍手状態と推定することを特徴とする請求項 2 6 記載の観客状況推定方法。

【請求項 3 0】 観客からの音声に応じた音声信号を生成し、前記音声信号に基づいて、前記観客からの音声の音量を検出し、前記音量に基づいて観客状況を推定することを特徴とする観客状況推定方法。

【請求項 3 1】 前記音声信号の音声レベルの分散を前記音量とすることを特徴とする請求項 3 0 記載の観客状況推定方法。

【請求項 3 2】 前記音声信号の音声レベルの分散を算出して、該分散の時間方向の平均値である時間マクロ音量を前記音量とすることを特徴とする請求項 3 0 記載の観客状況推定方法。

【請求項 3 3】 前記音量が所定レベルよりも大きいときには前記観客状況を笑い状態と推定することを特徴とする請求項 3 0 記載の観客状況推定方法。

【請求項 3 4】 観客からの音声に応じた音声信号を生成し、前記音声信号に基づいて、前記観客からの音声の周期性を検出し、前記音声の周期性に基づいて観客状況を推定することを特徴とする観客状況推定方法。

【請求項 3 5】 前記音量から自己相関最大位置を順次検出し、該自己相関



最大位置の分散を、前記音声の周期性とすることを特徴とする請求項 3 4 記載の観客状況推定方法。

【請求項 3 6】 前記音声信号に基づいて検出された前記観客からの音声の音量における低周波成分比を、前記音声の周期性とすることを特徴とする請求項 3 4 記載の観客状況推定方法。

【請求項 3 7】 前記音声の周期性が所定レベルよりも大きいときには前記観客状況を拍手状態と推定することを特徴とする請求項 3 4 記載の観客状況推定方法。

【請求項 3 8】 観客からの音声に応じた音声信号を生成し、前記音声信号のフーリエ変換を行い、該フーリエ変換結果の重心を検出し、前記重心に基づいて観客状況を推定することを特徴とする観客状況推定方法。

【請求項 3 9】 前記重心が所定レベルよりも大きいときには前記観客状況を手拍子状態あるいは拍手状態と推定することを特徴とする請求項 3 8 記載の観客状況推定方法。

【請求項 4 0】 観客からの音声に応じた音声信号を生成し、前記音声信号のフーリエ変換を行い、該フーリエ変換結果に基づいて高周波対低周波成分比を算出し、前記高周波対低周波成分比に基づいて観客状況を推定することを特徴とする観客状況推定方法。

【請求項 4 1】 前記高周波対低周波成分比によって、前記低周波成分に対する前記高周波成分の割合が所定レベルよりも大きいときには前記観客状況を手拍子状態あるいは拍手状態と推定することを特徴とする請求項 4 0 記載の観客状況推定方法。

【請求項 4 2】 観客を撮影した画像信号および／または前記観客からの音声に応じた音声信号を生成し、

前記画像信号に基づき、前記観客の動き量や動き周期性のいずれか 1 つあるいは両方を検出し、前記音声信号に基づき、前記観客のからの音声の音量や前記音声の周期性や前記音声の周波数成分に関する情報のいずれか 1 つあるいは複数を

検出し、

前記検出の結果に基づいて、観客状況を推定することを特徴とする観客状況推定方法。

【請求項 4 3】 コンピュータに、

観客を撮影した画像信号に基づき、前記観客の動き量や動き周期性のいずれか 1 つあるいは両方を検出し、および／または前記観客からの音声に応じた音声信号に基づき、前記観客のからの音声の音量や前記音声の周期性や前記音声の周波数成分に関する情報のいずれか 1 つあるいは複数を検出する手順と、

前記検出の結果に基づいて、観客状況を推定する手順とを実行させるための観客状況推定プログラム。

【発明の詳細な説明】

【0 0 0 1】

【発明の属する技術分野】

この発明は、観客状況推定装置と観客状況推定方法および観客状況推定プログラムに関する。詳しくは、観客を撮影した画像信号や観客からの音声に応じた音声信号に基づき、観客の動き量や動き周期性、音量、音声の周期性等が検出されて、この検出結果に基づき観客状況を推定するものである。

【0 0 0 2】

【従来の技術】

従来、映画や番組等のコンテンツを提供する場合、画像表示装置や音声出力装置等を用いて、予め作成されている画像信号や音声信号に基づいて画像表示や音声提示が行われている。

【0 0 0 3】

ここで、コンテンツの提供を効果的に行うためには、視聴者の状況を把握することが重要である。例えば、どのようなシーンで笑いや拍手が生じたを把握すれば、コンテンツの提供の際に、視聴者の反応状況に応じた効果音等を付加することで、より効果的なコンテンツ提供を行うことができる。また、コンテンツの製作者は、視聴者の反応状況を利用して、笑いが生じたシーンや拍手が生じたシーンを容易に判別することが可能となり、新たにコンテンツを制作する上で、視聴

者の反応状況を有効に活用することができる。また、コンテンツの制作段階で視聴者の反応状況を利用すれば、さらに効果的なコンテンツ制作を行うことも可能となる。

【 0 0 0 4 】

【発明が解決しようとする課題】

ところで、このように視聴者の反応状況を把握する場合、人が視聴者を観察して、どのようなシーンでどのような反応状況を示したかを判別しなければならないため、簡単に視聴者の反応状況を把握することができない。

【 0 0 0 5 】

そこで、この発明では、人が観客状況を観察しなくとも、観客状況を容易に推定できる観客状況推定装置および観客状況推定方法および観客状況推定プログラムを提供するものである。

【 0 0 0 6 】

【課題を解決するための手段】

この発明に係る観客状況推定装置は、観客を撮影した画像信号を生成する撮像手段と、前記画像信号に基づいて、前記観客の動き量を検出する動き量検出手段と、前記動き量に基づいて観客状況を推定する推定手段とを有するものである。

【 0 0 0 7 】

また、観客を撮影した画像信号を生成する撮像手段と、前記画像信号に基づいて、前記観客の動き周期性を検出する動き周期性検出手段と、前記観客の動き周期性に基づいて観客状況を推定する推定手段とを有するものである。

【 0 0 0 8 】

また、観客からの音声に応じた音声信号を生成する音声取得手段と、前記音声信号に基づいて、前記観客からの音声の音量を検出する音量検出手段と、前記音量に基づいて観客状況を推定する推定手段とを有するものである。

【 0 0 0 9 】

また、観客からの音声に応じた音声信号を生成する音声取得手段と、前記音声信号に基づいて、前記観客からの音声の周期性を検出する音声周期性検出手段と、前記音声の周期性に基づいて観客状況を推定する推定手段とを有するものであ

る。

【 0 0 1 0 】

また、観客からの音声に応じた音声信号を生成する音声取得手段と、前記音声信号のフーリエ変換を行い、該フーリエ変換結果の重心を検出するスペクトル検出手段と、前記重心に基づいて観客状況を推定する推定手段とを有するものである。

【 0 0 1 1 】

また、観客からの音声に応じた音声信号を生成する音声取得手段と、前記音声信号のフーリエ変換を行い、該フーリエ変換結果に基づいて高周波対低周波成分比を算出する成分比検出手段と、前記高周波対低周波成分比に基づいて観客状況を推定する推定手段とを有するものである。

【 0 0 1 2 】

さらに、観客を撮影した画像信号および／または前記観客からの音声に応じた音声信号を生成する入力手段と、前記画像信号に基づき、前記観客の動き量や動き周期性をいずれか1つあるいは両方を検出し、前記音声信号に基づき、前記観客のからの音声の音量や前記音声の周期性や前記音声の周波数成分に関する情報のいずれか1つあるいは複数を検出する特徴量検出手段と、前記特徴量検出手段での検出結果に基づいて、観客状況を推定する推定手段とを有するものである。

【 0 0 1 3 】

次に、この発明に係る観客状況推定方法は、観客を撮影して画像信号を生成し、前記画像信号に基づいて、前記観客の動き量を検出し、前記動き量に基づいて観客状況を推定するものである。

【 0 0 1 4 】

また、観客を撮影して画像信号を生成し、前記画像信号に基づいて、前記観客の動き周期性を検出し、前記観客の動き周期性に基づいて観客状況を推定するものである。

【 0 0 1 5 】

また、観客からの音声に応じた音声信号を生成し、前記音声信号に基づいて、前記観客からの音声の音量を検出し、前記音量に基づいて観客状況を推定するも

のである。

【 0 0 1 6 】

また、観客からの音声に応じた音声信号を生成し、前記音声信号に基づいて前記観客からの音声の周期性を検出し、前記音声の周期性に基づいて観客状況を推定するものである。

【 0 0 1 7 】

また、観客からの音声に応じた音声信号を生成し、前記音声信号のフーリエ変換を行い、該フーリエ変換結果の重心を検出し、前記重心に基づいて観客状況を推定するものである。

【 0 0 1 8 】

また、観客からの音声に応じた音声信号を生成し、前記音声信号のフーリエ変換を行い、該フーリエ変換結果に基づいて高周波対低周波成分比を算出し、前記高周波対低周波成分比に基づいて観客状況を推定するものである。

【 0 0 1 9 】

さらに、観客を撮影した画像信号および／または前記観客からの音声に応じた音声信号を生成し、前記画像信号に基づき、前記観客の動き量や動き周期性のいずれか1つあるいは両方を検出し、前記音声信号に基づき、前記観客のからの音声の音量や前記音声の周期性や前記音声の周波数成分に関する情報のいずれか1つあるいは複数を検出し、前記検出の結果に基づいて、観客状況を推定するものである。

【 0 0 2 0 】

さらに、この発明に係る観客状況推定プログラムは、コンピュータに、観客を撮影した画像信号に基づき、前記観客の動き量や動き周期性のいずれか1つあるいは両方を検出し、および／または前記観客からの音声に応じた音声信号に基づき、前記観客のからの音声の音量や前記音声の周期性や前記音声の周波数成分に関する情報のいずれか1つあるいは複数を検出する手順と、前記検出の結果に基づいて、観客状況を推定する手順とを実行させるものである。

【 0 0 2 1 】

この発明においては、観客を撮影した画像信号および／または観客からの音声

に応じた音声信号が生成されて、この画像信号や音声信号に基づき、例えば観客の動き量や動きの周期性、観客からの音声の音量や音声の周期性が検出されて、この検出結果に基づき観客が手拍子状態や拍手状態あるいは笑い状態等のいずれであるかが推定される。

#### 【 0 0 2 2 】

##### 【発明の実施の形態】

以下、図を参照しながら、この発明の実施の一形態について説明する。図 1 は、観客状況推定装置 1 0 の構成を示している。観客状況推定装置 1 0 の入力部 2 0 では、観客の状態をビデオカメラやマイクロフォンを用いて監視して監視情報を生成する。例えば、観客を撮影することにより、監視情報として撮影画像信号（以下「画像信号」という）S V を生成する。また、観客からの音声を取得することにより、監視情報として音声信号 S A の生成を行う。特徴量検出部 3 0 では、監視情報である画像信号 S V と音声信号 S A に基づいて複数の特徴量を検出する。推定部 4 0 では、検出された複数の特徴量に基づき、観客状態を推定して出力部 5 0 に供給する。出力部 5 0 では、推定部 4 0 での推定結果を出力する。

#### 【 0 0 2 3 】

図 2 は、入力部 2 0 の動作を説明するための図である。映画や番組等の画像は、視聴者 6 0 の前方に設けられた画像表示装置 6 1 で表示されると共に、音声はスピーカ 6 2 から出力される。入力部 2 0 は、撮像部 2 1 と音声取得部 2 2 で構成されている。撮像部 2 1 は、観客状況推定対象である視聴者 6 0 の前方に設けられて、例えば図 3 に示すような視聴者 6 0 側の撮影画像を示す画像信号 S V を生成する。音声取得部 2 2 は、視聴者 6 0 側からの音声を取得して音声信号 S A を生成する。このようにして生成された画像信号 S V と音声信号 S A は、特徴量検出部 3 0 に供給される。

#### 【 0 0 2 4 】

図 4 は、特徴量検出部 3 0 の構成を示している。この特徴量検出部 3 0 は、特徴量として例えば視聴者 6 0 側の動きの大きさや周期性および音声の音量や周期性を検出する。さらに、音声の周波数成分に関する特徴量を検出するものである。

## 【 0 0 2 5 】

特徴量検出部 3 0 の動き量検出部 3 1 では、動きの大きさを示す特徴量を検出する。動き周期性検出部 3 2 では、動きの周期性を示す特徴量を検出する。パワースペクトル検出部 3 4 では、音声の周波数成分に関する特徴量を検出する。また、音量検出部 3 5 では、音声の大きさを示す特徴量を検出すると共に、音声周期性検出部 3 6 では、音声の周期性を示す特徴量を検出する。

## 【 0 0 2 6 】

図 5 は、動き量検出部 3 1 の構成を示している。肌色領域抽出部 3 1 1 では、RGB 色空間や YIQ 色空間あるいは HSV 色空間等において、肌色と認識できる画素範囲を判別する。例えば、RGB 色空間では、肌色と認識する赤色信号の範囲と緑色信号の範囲と青色信号の範囲を示す閾値を設定して、画像信号 SV に基づいて生成した三原色信号の信号レベルが肌色の領域内であるか否かを画素毎に判別することで、肌色と認識できる画素範囲（以下「肌色領域」という）を抽出できる。

## 【 0 0 2 7 】

次に、動きベクトル計算部 3 1 2 では、肌色領域の動きベクトルを計算する。この動きベクトルの計算では、抽出された肌色領域をブロックに分割して、ブロック毎に動きベクトルを計算する。例えば、各視聴者の顔の部分や手の部分をそれぞれブロックとして、各ブロックについて次フレーム（または前フレーム）の画像とでブロックマッチングを行い、最もブロックの画像が一致したときの移動方向と移動量を動きベクトル MV とする。このようにして、式（1）に示すように 2 次元動きベクトル  $MV_i$  を検出する。

## 【 0 0 2 8 】

【数 1】

$$V_i = (u_i, v_i) \quad \cdots (1)$$

## 【 0 0 2 9 】

なお、式（1）において、「i」はブロック番号を示しており、ブロック数が例

例えば「 $m+1$ 」であるときには「 $i=0\sim m$ 」となる。また「 $u$ 」は水平方向の移動量、「 $v$ 」は垂直方向の移動量とする。さらに、動きベクトルの計算はブロックマッチング法に限られるものではなく、例えば代表点マッチング法や勾配法等の他の方法を用いるものとしても良い。

【0030】

平均動き量計算部313は、式(2)に基づき、フレーム内の各ブロックの動きベクトルを一定期間分だけ順次用いて、動きベクトルの大きさの移動平均を計算して平均動き量 $W_j$ とする。この算出した平均動き量 $W_j$ を特徴量 $CVa$ とする。なお、移動平均を算出する際に重み付けを行って平均動き量 $W_j$ を算出することもできる。

【0031】

【数2】

$$W_j = \sum_{i=0}^m \frac{\sqrt{u_i^2 + v_i^2}}{m} \quad \dots (2)$$

【0032】

図6は、特徴量 $CVa$ すなわちフレーム毎に算出した平均動き量 $W_j$ の推移を例示したものである。

【0033】

また、式(2)では、水平方向と垂直方向を足し合わせて平均動き量 $W_j$ を計算したが、水平方向あるいは垂直方向のいずれか一方に基づいて平均動き量 $W_j$ を計算することもできる。

【0034】

図7は、動き周期性検出部32の構成を示している。自己相関計算部321では、動き量検出部31で求めた平均動き量 $W_j$ を用いて、フレーム毎の自己相関係数 $RV$ を式(3)に基づいて計算する。

【0035】



【数 3】

$$RV = \frac{\sum_{j=0}^{n-1-k} W_j \cdot W(j+k)}{\sqrt{\sum_{j=0}^{n-1-k} W_j^2} \sqrt{\sum_{j=0}^{n-1-k} W(j+k)^2}} \quad \dots (3)$$

【0036】

図 8 は、式 (3) で示されている自己相関係数の算出動作を説明するためのものであり、説明を簡単とするため平均動き量  $W_j$  が図 8 A に示すように正弦波である場合を示している。ここで、「 $k=1$ 」のときには、図 8 B に示す最初の「0」フレームから  $(n-2)$  フレームまでの平均動き量  $W_j$  と、図 8 C に示す「1」フレームから  $(n-1)$  フレームまでの平均動き量  $W_{j+k}$  を用いて自己相関係数  $RV$  を算出する。同様に、「 $k=a$ 」のときには、図 8 D に示す最初の「0」フレームから  $(n-1-a)$  フレームまでの平均動き量  $W_j$  と、図 8 E に示す「 $a$ 」フレームから  $(n-1)$  フレームまでの平均動き量  $W_{j+k}$  を用いて自己相関係数  $RV$  を算出する。このように、自己相関係数  $RV$  の算出に用いる平均動き量を順次変更することで、図 8 F に示すように、「 $k=1$ 」から「 $k=a$ 」フレームまでの自己相関係数  $RV$  の推移を求めることができる。

【0037】

最大位置抽出部 322 は、算出された自己相関係数  $RV$  から、最大値となるフレーム位置を検出する。例えば、式 (3) に基づいて算出した自己相関係数  $RV$  の推移が図 9 に示す波形となったとき、フレーム位置  $PF_{mx}$  が自己相関最大位置  $MP$  となる。また、自己相関係数を算出する平均動き量  $W_j$  の位置を時間方向に順次シフトさせて同様な処理を行い、自己相関最大位置  $MP$  を順次検出すると、時間方向の自己相関最大位置  $MP$  の推移は例えば図 10 に示すものとなる。

【0038】

分散計算部 323 は、自己相関最大位置  $MP$  の分散を算出する。ここで分散  $V_p$  を算出する場合、検出する観客状況の周期性に応じて分散の算出に用いるフレ

ーム範囲を決定する。すなわち、周期性を検出するためには、少なくとも検出する周期以上の窓幅（フレーム範囲）の信号を用いなければ周期性を検出できない。また、窓幅を大きくすると、検出結果の精度を向上できると共に、周期変動が生じて正しい検出結果を得ることができる。しかし、窓幅が大きくなることにより、短時間の挙動の変化に対する応答が鈍くなると共に処理時間も長くなってしまう。このため、窓幅は検出する周期の数倍程度が好ましい。ここで、観客状況として視聴者 6 0 の例えば手拍子を推定するとき、手拍子の周期を 0.5 秒程度と考えて 2 秒程度を分散の算出に用いる窓幅とする。

【 0 0 3 9 】

ここで設定された窓幅、例えば p フレームに対して  $\pm (q/2)$  フレームを窓幅、この窓幅での自己相関最大位置の平均値を  $MP_{av}$ 、j フレーム目での自己相関最大位置を  $MP_j$  とすると、式 (4) に基づいて分散  $V_p$  を算出できる。

【 0 0 4 0 】

【数 4】

$$V_p = \frac{1}{q} \sum_{j=p-(q/2)}^{p+(q/2)} (MP_j - MP_{av})^2 \quad \dots (4)$$

【 0 0 4 1 】

このように、動き周期性検出部 3 2 は、平均動き量の自己相関最大位置  $MP$  の分散  $V_p$  を算出して、この分散  $V_p$  を特徴量  $CV_b$  とする。この特徴量  $CV_b$  の値が大きければ周期の変動が大きく、特徴量  $CV_b$  の値が小さければ周期の変動が小さい。なお、特徴量  $CV_b$  は、信号の周期の変動具合を示すものであれば良く、例えば自己相関最大位置  $MP$  の標準偏差等を用いることもできる。なお、図 1 1 は式 (4) に基づいて検出した特徴量  $CV_b$  の推移を例示したものである。

【 0 0 4 2 】

次に音声信号  $SA$  に基づいた特徴量について説明する。フィルタ部 3 3 は、音声信号  $SA$  の帯域制限を行い、パワースペクトル検出部 3 4 や音量検出部 3 5 および音声周期性検出部 3 6 で特徴量を生成するために適した周波数帯域の成分を抽出する。なお、この周波数帯域の以外の成分が音声信号  $SA$  に含まれていない

とき、あるいはこの周波数帯域の以外の成分が少なく特徴量を正しく生成できる  
ときには、フィルタ部 3 3 を設けなくても良いことは勿論である。

## 【 0 0 4 3 】

パワースペクトル検出部 3 4 では、フィルタ部 3 3 を介して供給された音声信号 S A を用いて、検出する観客状況に応じて設定した時間分の音声信号 S A に基づき重心を算出する。例えば、視聴者 6 0 の「笑い」状態を検出する場合には、重心を算出する際に笑い声の周波数成分が含まれるように時間長を設定する。図 1 2 は、パワースペクトル検出部 3 4 の構成を示している。

## 【 0 0 4 4 】

フーリエ変換部 3 4 1 は、設定した時間分の音声信号 S A のフーリエ変換を行い、周波数毎のパワーを検出する。例えば図 1 3 に示す音声信号 S A のフーリエ変換結果は図 1 4 に示すものとなる。

## 【 0 0 4 5 】

重心計算部 3 4 2 では、周波数  $f_i$  に対してパワー  $PW_i$  の重みを付けた平均、すなわち重心  $CA$  を求める。この重心  $CA$  は、重心  $CA$  を求める際の周波数範囲を「 $f_0 \sim f_g$ 」としたとき、式 (5) に基づいて算出できる。この算出した重心  $CA$  を特徴量  $CV_c$  とする。図 1 5 は、特徴量  $CV_c$  の推移を例示したものである。

## 【 0 0 4 6 】

【数 5】

$$CA = \sum_{i=0}^g f_i \cdot PW_i / g \quad \dots (5)$$

## 【 0 0 4 7 】

音量検出部 3 5 では、フィルタ部 3 3 を介して供給された音声信号 S A を用いて、予め設定した区間、例えば 1 フレーム期間分の音声レベルの分散を算出する。なお分散の算出は、上述の式 (4) と同様にして算出できる。この算出した分散は、音量を示す特徴量  $CV_d$  とする。また、分散を算出する際には、音声信号

を用いる 1 フレーム期間を順次移動させて、移動平均の算出と同様にして分散を算出する。図 1 6 は、特徴量 C V d の推移を例示したものである。

【 0 0 4 8 】

音声周期性検出部 3 6 では、画像の場合と同様にして、自己相関最大位置の分散を算出して特徴量 C V e とする。図 1 7 は音声周期性検出部 3 6 の構成を示している。

【 0 0 4 9 】

自己相関計算部 3 6 1 では、フィルタ部 3 3 を介して供給された音声信号 S A の例えば所定区間分を用いて、この音声信号 S A の信号レベル  $a_j$  から式 ( 6 ) に基づき自己相関係数 R A を算出する。なお、所定区間は、上述の自己相関係数 R V の算出と同様に、検出する音声の周期よりも長く設定する。

【 0 0 5 0 】

【数 6】

$$R A = \frac{\sum_{j=0}^{r-1-s} a_j \cdot a(j+s)}{\sqrt{\sum_{j=0}^{r-1-s} a_j^2} \sqrt{\sum_{j=0}^{r-1-s} a(j+s)^2}} \quad \dots (6)$$

【 0 0 5 1 】

最大位置抽出部 3 6 2 は、算出された自己相関係数 R A から、最大となる位置を、検出する観客状況に応じて設定された時間範囲毎に検出する。このように自己相関係数 R A が最大となる位置を検出することで、自己相関最大位置 M Q は例えば図 1 8 に示すものとなる。

【 0 0 5 2 】

分散計算部 3 6 3 は、自己相関最大位置 M Q の分散を算出する。ここで自己相関最大位置の分散の算出は、上述の式 ( 4 ) と同様な演算処理を行うことで算出できる。この算出した分散を特徴量 C V e とする。なお、図 1 9 は特徴量 C V e の推移を例示したものである。

## 【 0 0 5 3 】

このように、音声周期性検出部 3 6 は、音声の自己相関最大位置の分散を特徴量  $CV_e$  とする。この特徴量  $CV_e$  の値が大きければ周期の変動が大きく、特徴量  $CV_e$  の値が小さければ周期の変動が小さい。なお、特徴量  $CV_e$  は、音声信号の周期の変動具合を示すものであれば良く、例えば自己相関最大位置の標準偏差等を用いることもできる。

## 【 0 0 5 4 】

推定部 4 0 は、特徴量検出部 3 0 で検出された複数の特徴量を用いて、視聴者 6 0 の状況を推定する。図 2 0 は、特徴量検出部 3 0 で検出された特徴量  $CV_a$  ,  $CV_b$  ,  $CV_c$  ,  $CV_d$  ,  $CV_e$  を用いて、視聴者 6 0 の状況を推定する場合を示している。動きの大きさを示す特徴量  $CV_a$  やパワースペクトルの重心を示す特徴量  $CV_c$  が、基準レベルよりも大きいときには、例えば「手拍子」あるいは「拍手」のいずれかの状態と推定する。また、基準レベルよりも小さいときには、「笑い」あるいは「その他の動作」のいずれかの状態と推定する。なお、基準レベルは、各特徴量に応じて予め設定されるものであり、以下の場合も同様である。

## 【 0 0 5 5 】

例えば、図 2 1 A に示すように、特徴量  $CV_a$  が基準レベル  $L_{v1}$  よりも大きいときには「手拍子」あるいは「拍手」のいずれかの状態と推定すると共に、基準レベル  $L_{v1}$  以下であるときには、「笑い」あるいは「その他の動作」のいずれかの状態と推定する。

## 【 0 0 5 6 】

また、図 2 1 B に示すように、特徴量  $CV_c$  が基準レベル  $L_{a1}$  よりも大きいときには「手拍子」あるいは「拍手」のいずれかの状態と推定すると共に、基準レベル  $L_{a1}$  以下であるときには「笑い」あるいは「その他の動作」のいずれかの状態と推定する。

## 【 0 0 5 7 】

次に、動きの周期性を示す特徴量  $CV_b$  や音声の周期性を示す特徴量  $CV_e$  が基準レベルよりも大きいときには「拍手」の状態と推定すると共に、基準レベルよ

りも小さいときには「手拍子」の状態と推定する。

【 0 0 5 8 】

例えば、図 2 2 A に示すように、特徴量  $CVb$  が基準レベル  $Lv2$  よりも大きいときには「拍手」の状態と推定すると共に、基準レベル  $Lv2$  以下であるときには「手拍子」の状態と推定する。また、図 2 2 B に示すように、特徴量  $CVe$  が基準レベル  $La2$  よりも大きいときには「拍手」の状態と推定すると共に、基準レベル  $La2$  以下であるときには「手拍子」の状態と推定する。

【 0 0 5 9 】

さらに、音声の周期性を示す特徴量  $CVd$  が基準レベルよりも大きいときには「笑い」の状態と推定すると共に、基準レベルよりも小さいときには「その他の動作」の状態と推定する。例えば、図 2 2 C に示すように、特徴量  $CVd$  が基準レベル  $La3$  よりも大きいときには「笑い」の状態と推定すると共に、基準レベル  $La3$  以下であるときには「その他の動作」の状態と推定する。

【 0 0 6 0 】

また、動きの周期性を示す特徴量  $CVb$  の推定結果と音声の周期性を示す特徴量  $CVe$  の推定結果の論理積あるいは論理和、例えば「拍手」を論理「0」、「手拍子」を論理「1」として論理演算を行い、演算結果を最終推定結果としても良い。この場合には、動きの周期性に基づく推定結果と音声の周期性に基づく推定結果が異なる場合が生じて、1つの推定結果を得ることができる。

【 0 0 6 1 】

さらに、視聴者 60 の状況の推定に用いる特徴量をコンテンツ提供状況に応じて選択して用いるものとしても良い。すなわち、視聴者 60 の動きを良好に取得できると共に音声の取得が困難なときには、動きの大きさや周期性を示す特徴量を主として用いることで観客状況を推定する。また、視聴者 60 の動きの取得が困難であると共に音声を良好に取得できるときには、音声の音量や周期性を示す特徴量を主として用いることで観客状況を推定する。例えば、図 2 3 に示すように、コンサート等では、○印で示すように、動き大きさや音声の周期性および音声の音量を示す特徴量  $CVa$ 、 $CVe$ 、 $CVd$  を用いるものとする。また、映画館では、館内が暗いことから、パワースペクトルや音声の周期性および音声の音量

を示す特徴量  $CV_c$ ,  $CV_e$ ,  $CV_d$  を用いるものとする。さらに、スポーツ中継では、コンサート等に比べて音声の周期性が少ないことから、動きの大きさや動きの周期性および音声の音量を示す特徴量  $CV_a$ ,  $CV_b$ ,  $CV_d$  を用いるものとする。このように、観客状況の推定に用いる特徴量をコンテンツ提供状況に応じて選択することで、より適正な観客状況の推定を行うことができる。

## 【 0 0 6 2 】

また、複数の特徴量を統合して、統合結果と閾値を比較して観客状況の推定を行うものとしても良い。例えば、動きの大きさや音声の音量等を平均して平均値を算出すると共に、この平均値と閾値を比較して状況推定を行う。あるいは、コンテンツの提供状況に応じて動きの大きさや音声の音量等に重み付けを行い、この重み付けがなされた特徴量の平均値を用いて状況推定を行うこともできる。さらに、特徴量の加算値や線形和すなわち重み付けがなされた特徴量の加算値を算出して、この加算値と閾値を比較した比較結果に基づき状況推定を行うものとしても良い。

## 【 0 0 6 3 】

出力部 50 では、表示器を用いて推定結果を表示する。あるいは、提供しているコンテンツが記録媒体に記録されているコンテンツであるとき、あるいは提供しているコンテンツを記録媒体に記録しているとき、この推定結果を合わせて記録媒体に記録すれば、視聴者が所望の状態となった部分を容易に検索できる。

## 【 0 0 6 4 】

このように、人が視聴者を観察しなくとも、視聴者 60 側の撮影画像を示す画像信号  $SV$  や視聴者 60 側からの音声を取得して生成した音声信号  $SA$  に基づき、視聴者 60 側の動きや音声の大きさや周期性等を判別して、視聴者 60 の状況を自動的に推定できる。

## 【 0 0 6 5 】

また、上述の観客状況の推定は、ハードウェアだけでなくソフトウェアで実現するものとしても良い。この場合の構成を図 24 に示す。コンピュータは、図 24 に示すように CPU (Central Processing Unit) 701 を内蔵しており、この CPU 701 にはバス 720 を介して ROM 702, RAM 703, ハード・デ

ィスク・ドライブ 7 0 4，入出力インタフェース 7 0 5 が接続されている。さらに、入出力インタフェース 7 0 5 には操作入力部 7 1 1 や記録媒体ドライブ 7 1 2，通信部 7 1 3，信号入力部 7 1 4，信号出力部 7 1 5，表示部 7 1 6 が接続されている。

## 【 0 0 6 6 】

外部装置から命令が入力されたり、キーボードやマウス等の操作手段あるいはマイク等の音声入力手段等を用いて構成された操作入力部 7 1 1 から命令が入力されたりすると、この命令が入出力インタフェース 7 0 5 を介して CPU 7 0 1 に供給される。

## 【 0 0 6 7 】

CPU 7 0 1 は、ROM 7 0 2 や RAM 7 0 3 あるいはハード・ディスク・ドライブ 7 0 4 に記憶されているプログラムを実行して、供給された命令に応じた処理を行う。さらに、ROM 7 0 2 や RAM 7 0 3 あるいはハード・ディスク・ドライブ 7 0 4 には、上述の推定処理をコンピュータで実行させるための観客状況推定プログラムを予め記憶させて、信号入力部 7 1 4 に入力された信号に基づいて観客状況を推定して、推定結果を信号出力部 7 1 5 から出力する。また、記録媒体に観客状況推定プログラムを記録しておくものとし、記録媒体ドライブ 7 1 2 によって、観客状況推定プログラムを記録媒体に記録しあるいは記録媒体に記録されている観客状況推定プログラムを再生してコンピュータで実行するものとしても良い。さらに、通信部 7 1 3 によって、観客状況推定プログラムを有線あるいは無線の伝送路を介して送信あるいは受信するものとし、受信した観客状況推定プログラムをコンピュータで実行するものとしても良い。

## 【 0 0 6 8 】

次に、観客状況推定プログラムについて説明する。図 2 5 は、観客状況推定処理の全体を示すフローチャートを示している。

## 【 0 0 6 9 】

ステップ S T 1 では、画像信号と音声信号の取り込みを行う。例えば、信号入力部 7 1 4 に供給された画像信号 S V と音声信号 S A を用いて、観客状況の推定を開始する。



## 【 0 0 7 0 】

ステップ S T 2 では、特徴量の検出を行う。図 2 6 は特徴量 C V a の検出動作を示すフローチャートである。また図 2 7 ～ 図 3 0 は、特徴量 C V b, C V c, C V d, C V e の各検出動作を示している。

## 【 0 0 7 1 】

図 2 6 のステップ S T 1 1 では、画像信号 S V に基づき肌色領域を抽出する。ステップ S T 1 2 では、ステップ S T 1 1 で抽出した肌色領域の動きベクトルをブロック毎に算出する。ステップ S T 1 3 ではステップ S T 1 2 で算出したブロック毎の動きベクトルに基づき、各ブロックの動き量をフレーム毎に平均して平均動き量を算出する。さらに、算出した平均動き量を特徴量 C V a とする。

## 【 0 0 7 2 】

図 2 7 のステップ S T 2 1 は、ステップ S T 1 3 で算出した平均動き量を用いて自己相関係数を算出する。ステップ S T 2 2 では、予め設定された所定区間（k フレームまで）分だけ完了した否かを判別して、完了していないときにはステップ S T 2 1 に戻る。また完了したときにはステップ S T 2 3 に進む。

## 【 0 0 7 3 】

ステップ S T 2 3 では、所定区間分の自己相関係数から最大となる位置を判別してステップ S T 2 4 に進む。ステップ S T 2 4 では、自己相関最大位置の抽出が検出する動きの周期性に応じて設定された窓幅分だけ行われたか否かを判別する。このステップ S T 2 4 で、窓幅分の抽出が完了していないときにはステップ S T 2 1 に戻る。また、抽出が完了したときにはステップ S T 2 5 に進む。

## 【 0 0 7 4 】

ステップ S T 2 5 は、抽出した自己相関最大位置を用いて分散を算出する。この算出した分散を特徴量 C V b とする。

## 【 0 0 7 5 】

図 2 8 のステップ S T 3 1 では、検出する観客状況に応じて設定した時間分の音声信号 S A のフーリエ変換を行い、周波数毎のパワーを検出する。ステップ S T 3 2 では、重心を算出して周波数に対してパワーの重みを付けた平均、すなわち重心 C A を求める。この重心 C A を特徴量 C V c とする。

## 【 0 0 7 6 】

図 2 9 のステップ S T 4 1 では、予め設定した区間分、例えばフレーム分の音声信号 S A を抽出してステップ S T 4 2 に進む。ステップ S T 4 2 では抽出した音声信号 S A の分散を算出して特徴量 C V d とする。

## 【 0 0 7 7 】

図 3 0 のステップ S T 5 1 では、音声信号 S A の信号レベルの自己相関係数を算出する。ステップ S T 5 2 では、自己相関の算出を予め定められた所定区間分だけ完了した否かを判別して、完了していないときにはステップ S T 5 1 に戻る。また完了したときにはステップ S T 5 3 に進む。

## 【 0 0 7 8 】

ステップ S T 5 3 では、所定区間分の自己相関係数から最大となる位置を判別してステップ S T 5 4 に進む。ステップ S T 5 4 では、自己相関最大位置の抽出が検出する音声の周期性に応じて設定された窓幅分だけ行われたか否かを判別する。このステップ S T 5 4 で、窓幅分の抽出が完了していないときにはステップ S T 5 1 に戻る。また、抽出が完了したときにはステップ S T 5 5 に進む。

## 【 0 0 7 9 】

ステップ S T 5 5 は、抽出した自己相関最大位置を用いて分散を算出する。この算出した分散を特徴量 C V e とする。

## 【 0 0 8 0 】

このように、動き量や動きの周期性、音声の音量や周期性、パワースペクトルの重心を特徴量として検出する。図 2 5 のステップ S T 3 では、ステップ S T 2 で検出した複数の特徴量に基づいて、観客状況を推定する。図 3 1 は、観客状況推定処理を示すフローチャートである。

## 【 0 0 8 1 】

ステップ S T 6 1 では、平均動き量を示す特徴量 C V a あるいはパワースペクトルの重心を示す特徴量 C V c の何れかを選択してステップ S T 6 2 に進む。

## 【 0 0 8 2 】

ステップ S T 6 2 では、ステップ S T 6 1 で特徴量 C V a が選択されたか否かを判別する。ここで、特徴量 C V a が選択されたときにはステップ S T 6 3 に進

む。また特徴量C V cが選択されたときにはステップS T 7 3に進む。

【 0 0 8 3 】

ステップS T 6 3では、特徴量C V aが基準レベルL v1よりも大きいかな否かを判別する。ここで、特徴量C V aが基準レベルL v1よりも大きいときにはステップS T 6 4に進み、大きくないときにはステップS T 7 4に進む。

【 0 0 8 4 】

ステップS T 6 4では、観客状況が「手拍子」あるいは「拍手」のいずれかの状態であると推定してステップS T 6 4に進む。

【 0 0 8 5 】

ステップS T 6 5では、特徴量C V bあるいは特徴量C V eのいずれかを選択してステップS T 6 6に進む。ステップS T 6 6では、ステップS T 6 5で特徴量C V bが選択されたときステップS T 6 7に進む。また特徴量C V eが選択されたときにはステップS T 7 0に進む。

【 0 0 8 6 】

ステップS T 6 7では、特徴量C V bが基準レベルL v2よりも大きいかな否かを判別する。ここで、特徴量C V bが基準レベルL v2よりも大きいときにはステップS T 6 8に進み、観客状況は「拍手」の状態であると推定する。また、特徴量C V bが基準レベルL v2よりも大きくないときにはステップS T 6 9に進み、観客状況は「手拍子」の状態であると推定する。

【 0 0 8 7 】

ステップS T 6 5で特徴量C V eが選択されて、ステップS T 6 6からステップS T 7 0に進むと、ステップS T 7 0では、特徴量C V eが基準レベルL a2よりも大きいかな否かを判別する。ここで、特徴量C V eが基準レベルL a2よりも大きいときにはステップS T 7 1に進み、観客状況は「拍手」の状態であると推定する。また、特徴量C V eが基準レベルL a2よりも大きくないときにはステップS T 7 2に進み、観客状況は「手拍子」の状態であると推定する。

【 0 0 8 8 】

ステップS T 6 1で特徴量C V cが選択されて、ステップS T 6 2からステップS T 7 3に進むと、ステップS T 7 3では、特徴量C V cが基準レベルL a1よ

りも大きいかな否かを判別する。ここで、特徴量C V cが基準レベルL a1よりも大きいときにはステップS T 6 4に進む。また、特徴量C V cが基準レベルL a1よりも大きくないときにはステップS T 7 4に進む。

#### 【 0 0 8 9 】

ステップS T 6 3あるいはステップS T 7 3からステップS T 7 4に進むと、ステップS T 7 4では、観客状況が「笑い」あるいは「その他の動作」のいずれかの状態であると推定してステップS T 7 5に進む。

#### 【 0 0 9 0 】

ステップS T 7 5では、特徴量C V dが基準レベルL a3よりも大きいかな否かを判別する。ここで、特徴量C V dが基準レベルL a3よりも大きいときにはステップS T 7 6に進み、観客状況は「笑い」の状態であると推定する。また、特徴量C V dが基準レベルL a3よりも大きくないときにはステップS T 7 7に進み、観客状況は「その他の動作」の状態であると推定する。

#### 【 0 0 9 1 】

このようにして観客状況が推定されると図2 5のステップS T 4では、この推定結果を示す推定結果信号を信号出力部7 1 5から出力する。あるいは、表示部7 1 6に推定結果を表示する。また、記録媒体にコンテンツが記録されており、このコンテンツを再生して信号出力部7 1 5からコンテンツの画像信号や音声信号を出力する場合には、このコンテンツの再生位置と推定結果を関係付けて記録媒体に記録し、あるいは記憶手段に記憶させるものとしても良い。このようにすれば、コンテンツの内容と観客状況の関係を容易に確認できる。

#### 【 0 0 9 2 】

以上のように観客状況推定プログラムを実行することで、ソフトウェアに基づいても観客状況を推定することができる。

#### 【 0 0 9 3 】

ところで、上述の実施の形態では、動きの大きさを示す特徴量C V a、動きの周期性を示す特徴量C V b、パワースペクトルの重心を示す特徴量C V c、音声の音量を示す特徴量C V d、音声の周期性を示す特徴量C V eを用いて、視聴者6 0の状況を推定する場合を示したが、特徴量はこれらに限られるものではない。

## 【 0 0 9 4 】

図 3 2 は、特徴量検出部の他の構成を示している。この特徴量検出部 8 0 は、特徴量として例えば視聴者 6 0 側の一定時間内における動き量、動きの周期性、一定時間内における音量、音声の周期性を検出する。さらに、音声の周波数成分に関する特徴量を検出するものである。ここで、動き量や音量の検出は、上述の特徴量検出部 3 0 と同様に、1 フレーム毎の画像信号や 1 フレームに対応する期間毎の音声信号を用いて、一定期間内における動き量や音量の検出を行う。また、動きや音声の周期性の検出は、低域フィルタによる帯域制限やフーリエ変換によって低周波成分の検出を行い、検出された低周波成分に基づいて算出する。さらに、音声の周波数成分に関する特徴量の検出は、音声レベルのフーリエ変換結果を用いて行う。

## 【 0 0 9 5 】

特徴量検出部 8 0 の動き量検出部 8 1 では、平均動き量を検出する。動き周期性検出部 8 2 では、動きの周波数成分に基づいて周期性を検出する。音量検出部 8 4 では、音量を検出すると共に、音声周期性検出部 8 5 では、音声の周波数成分に基づいて周期性を検出する。高周波対低周波成分比算出部 8 6 では、音声の周波数成分に関する特徴量として音声の高周波対低周波成分比を算出する。

## 【 0 0 9 6 】

図 3 3 は動き量検出部 8 1 の構成を示している。なお、図 5 に示す動き量検出部 3 1 と対応する部分については同一符号を付している。肌色領域抽出部 3 1 1 は肌色領域を抽出すると共に、動きベクトル計算部 3 1 2 は肌色領域の動きベクトルを計算する。平均動き量計算部 3 1 3 は、式 (2) に基づき、フレーム毎に動きベクトルの大きさの平均を計算して、平均動き量  $W_j$  とする。この算出した平均動き量  $W_j$  を時間平均計算部 8 1 1 と動き周期性検出部 8 2 に供給する。

## 【 0 0 9 7 】

時間平均計算部 8 1 1 では、平均動き量  $W_j$  を時間方向に平均して時間マクロ動き量  $W_U$  を算出する。この時間マクロ動き量  $W_U$  の算出は、移動平均や重み付け移動平均の算出と同様に行う。例えば  $n$  フレーム分の平均動き量  $W_j$  を用いて時間マクロ動き量  $W_U$  を算出する場合、 $a$  フレームから  $(a + n)$  フレームまで

の平均動き量を用いて時間マクロ動き量 $WU_a$ を算出する。次に、 $(a + 1)$  フレームから  $(a + n + 1)$  フレームまでの平均動き量 $W_j$ を平均して時間マクロ動き量 $WU_{(a+1)}$ を算出する。以下同様にして、フレーム位置を順次移動させて時間マクロ動き量 $WU$ を順次算出すると共に、この時間マクロ動き量を特徴量 $CW_a$ とする。また、重み付け移動平均を算出する場合には、各フレームの平均動き量に重み付けを行ってから平均して時間マクロ動き量を算出する。さらにフレーム位置を順次移動させて時間マクロ動き量を算出して特徴量 $CW_a$ とする。この時間マクロ動き量の算出の際に平均するフレーム数は、推定する観客状況に応じて設定する。例えば「手拍子」や「拍手」「笑い」等の状態を推定する場合には、1～2秒程度のフレーム分の平均動き量 $W_j$ を用いて時間マクロ動き量 $WU$ を算出する。なお、図34Aおよび図34Bは、時間マクロ動き量 $WU$ である特徴量 $CW_a$ の時間方向の推移を例示している。

## 【0098】

動き周期性検出部82は、動きベクトルの大きさの平均を示す平均動き量、すなわち検出する動きの周期性に応じて設定された窓幅内の平均動き量 $W_j$ についてフーリエ変換を行う。このフーリエ変換結果に基づき、平均動き量 $W_j$ の低周波成分比、すなわちフーリエ変換結果に基づいて、全パワーに対する低周波成分のパワーの比を算出する。ここで、低周波成分の範囲は、検出する状態の周期性に応じて設定する。すなわち、検出する状態の周波数が「 $f_a$ 」で、他の状態の周波数が「 $f_b$ 」である場合、周波数「 $f_a$ 」と周波数「 $f_b$ 」との間に閾値を設定して、この閾値以下の成分を低周波成分とする。ここで低周波成分が大きいときには検出する状態であり、低周波成分が少ないときには他の状態であると推定できる。例えば、手拍子の周波数が2Hz程度で、拍手が5Hz程度あるときには、この2つの状態の周期が含まれるように窓幅を設定すると共に閾値を2Hzと5Hzの間に設定して、低周波成分を決定する。このようにして低周波成分比 $EV_a$ を算出すると共に、この低周波成分比 $EV_a$ を特徴量 $CW_b$ とする。また、直流成分を除いて低周波成分比 $EV_a$ を算出することで、周期性をより明確化できる。

## 【0099】

さらに、動き周期性検出部 8 2 は、低域フィルタを用いて時間マクロ動き量  $W_U$  から低域成分を抽出して、この低域成分と低域フィルタによる帯域制限前の時間マクロ動き量  $W_U$  との比を低周波成分比として求め、この低周波成分比を特徴量  $C_{Wb}$  としても良い。さらに、帯域フィルタを用いて、直流成分を除くものとするれば、更に精度良く周期性を検出できる。

## 【 0 1 0 0 】

図 3 5 A は周波数の低い状態であるときにフーリエ変換結果を示しており、図 3 5 B は周波数の低い状態でないときにフーリエ変換結果を示している。ここで、閾値を周波数  $L_{Fa}$  に設定すれば、低周波成分比  $E_{Va}$  に基づき、周波数の低い状態であるか否かを判別できる。また、図 3 6 は特徴量  $C_{Wb}$  の推移を例示したものである。

## 【 0 1 0 1 】

図 3 7 は、音量検出部 8 4 の構成を示している。音量検出部 8 4 の分散計算部 8 4 1 では、上述の音量検出部 3 5 と同様にして、予め設定した区間、例えばフレーム毎に音声レベルの分散を算出する。時間平均計算部 8 4 2 では、時間平均計算部 8 1 1 での時間マクロ動き量  $W_U$  の算出と同様な処理を行い、予め設定した区間、例えばフレーム毎の分散を、推定する観客状況に応じた所定区間毎に順次用いて移動平均あるいは重み付け移動平均の算出を行い、この算出された移動平均を時間マクロ音量  $W_T$  とする。さらに、この時間マクロ音量  $W_T$  を特徴量  $C_{Wc}$  とする。例えば図 3 8 A および図 3 8 B は、特徴量  $C_{Wc}$  の推移を例示したものである。

## 【 0 1 0 2 】

このように、フレーム毎に算出した音声レベルの分散を平均化することで、例えば単独フレームの音声レベルが、視聴内容や環境によって変動しても、変動の少ない安定した特徴量を得ることができる。

## 【 0 1 0 3 】

音声周期性検出部 8 5 では、音量検出部 8 4 で算出した時間マクロ音量  $W_T$  のフーリエ変換を行い、このフーリエ変換結果に基づき、全パワーに対する低周波成分のパワーの比を算出する。ここで、低周波成分の範囲は、動き周期性の場合

と同様に、検出する音声の周期性に応じて設定する。ここで、低周波成分が大きいときには検出する状態であり、低周波成分が少ないときには他の状態であると推定できる。この音声周期性検出部 8 5 で算出した低周波成分比を特徴量  $CWd$  とする。

#### 【 0 1 0 4 】

また、音声周期性検出部 8 5 では、低域フィルタを用いて時間マクロ音量  $WT$  から低域成分を抽出して、この低域成分と低域フィルタによる帯域制限前の時間マクロ音量  $WT$  との比を低周波成分比として求め、この低周波成分比を特徴量  $CWd$  としても良い。図 3 9 は、特徴量  $CWd$  の推移を例示したものである。なお、帯域フィルタを用いて、直流成分を除くものとするれば、更に精度良く周期性を検出できる。

#### 【 0 1 0 5 】

高周波対低周波成分比計算部 8 6 では、フィルタ部 3 3 を介して供給された音声信号  $SA$  を例えば推定する観客状況に応じた所定区間分だけ用いてフーリエ変換を行うことで所定区間分のパワーを高周波成分と低周波成分に区分して、高周波対低周波成分比を算出する。さらに、この高周波対低周波成分比を特徴量  $CWe$  とする。

#### 【 0 1 0 6 】

ここで、例えば「拍手」と「笑い」の状態での音声では、「笑い」よりも「拍手」の状態で高い周波数成分が多くなる。このため、高周波対低周波成分比に基づき観客状況を推定することが可能となる。

#### 【 0 1 0 7 】

なお、高周波対低周波成分比の算出では、高域フィルタによって音声信号  $SA$  から高周波成分を分離すると共に、低域フィルタを用いて低周波成分を分離して、得られた高周波成分と低周波成分から高周波対低周波成分比を算出することもできる。図 4 0 は特徴量  $CWe$  の推移を例示したものである。

#### 【 0 1 0 8 】

推定部 9 0 は、特徴量  $CWa$ ,  $CWb$ ,  $CWc$ ,  $CWd$ ,  $CWe$  を用いて、視聴者 6 0 の状態を推定する。図 4 1 は推定動作を示している。動き量を示す特徴量  $C$



Waや音量の大きさを示す特徴量CWcが、第1の基準レベルよりも大きいときには、例えば「手拍子」あるいは「拍手」のいずれかの状態と推定する。また、第1の基準レベルよりも低い第2の基準レベルよりも小さいときには、「その他の動作」の状態と推定する。また、第1の基準レベルと第2の基準レベルの範囲内であるときには「笑い」の状態と判断する。

## 【0109】

例えば図42Aに示すように、特徴量CWaが基準レベルLv12よりも大きい部分では、「手拍子」あるいは「拍手」のいずれかの状態と推定する。また図42Bに示すように、特徴量CWaが基準レベルLv11よりも小さい部分では、「その他の動作」の状態と推定すると共に、基準レベルLv11～Lv12の範囲は、「笑い」の状態と推定する。

## 【0110】

また、図43Aに示すように、特徴量CWcが基準レベルLa12よりも大きい部分では、「手拍子」あるいは「拍手」のいずれかの状態と推定する。また図43Bに示すように、特徴量CWcが基準レベルLa11よりも小さい部分では、「その他の動作」の状態と推定すると共に、基準レベルLa11～La12の範囲は、「笑い」の状態と推定する。

## 【0111】

次に、平均動き量の低周波成分比を示す特徴量CWbや音量の低周波成分比を示す特徴量CWdが基準レベルよりも大きいときには「手拍子」の状態と推定すると共に、基準レベルよりも小さいときには「拍手」の状態と推定する。

## 【0112】

例えば図44Aに示すように、特徴量CWbが基準レベルLv13よりも大きい部分では「手拍子」、基準レベルLv13以下の部分では「拍手」の状態と推定する。また図44Bに示すように、特徴量CWdが基準レベルLa13よりもより大きい部分では「手拍子」、基準レベルLa13以下の部分では「拍手」の状態と推定する。

## 【0113】

また高周波対低周波成分比を示す特徴量CWeを用いる場合には、図45に示

す推定動作を行う。動きの大きさを示す特徴量  $CW_a$  や音量の大きさを示す特徴量  $CV_c$  が、基準レベルよりも大きいときには、例えば「手拍子」あるいは「拍手」または「笑い」のいずれかの状態と推定する。また、基準レベルよりも小さいときには、「その他の動作」の状態と推定する。次に、音声の高周波対低周波成分比を示す特徴量  $CW_e$  が基準レベルよりも大きいときには「手拍子」あるいは「拍手」の状態と推定すると共に、基準レベルよりも小さいときには「笑い」の状態と推定する。例えば、図 4 6 に示すように、特徴量  $CW_e$  が基準レベル  $La_{14}$  よりも大きいときには「手拍子」あるいは「拍手」の状態と推定すると共に、基準レベル  $La_{14}$  よりも小さいときには「笑い」の状態と推定する。

## 【 0 1 1 4 】

動きの低周波成分比を示す特徴量  $CW_b$  や音量の低周波成分比を示す特徴量  $CW_d$  が基準レベルよりも大きいときには「手拍子」の状態と推定すると共に、基準レベルよりも小さいときには「拍手」の状態と推定する。さらに、音声の周期性を示す特徴量  $CV_d$  が基準レベルよりも大きいときには「笑い」の状態と推定すると共に、基準レベルよりも小さいときには「その他の動作」の状態と推定する。

## 【 0 1 1 5 】

また、上述したように、動きの低周波成分比を示す特徴量  $CW_b$  の推定結果と音量の低周波成分比を示す特徴量  $CW_d$  の推定結果の論理演算を行い、演算結果を最終推定結果としても良い。さらに、視聴者 6 0 の状態の推定に用いる特徴量をコンテンツ提供状況に応じて選択して用いるものとしても良いことは勿論である。

## 【 0 1 1 6 】

次に、動きや音声の周波数成分に基づいて特徴量を検出する場合の観客状況推定プログラムについて説明する。なお、この場合、図 2 5 のステップ  $ST_3$  に示す特徴量の検出とステップ  $ST_4$  の推定が異なる処理となり、他の処理は同一であることから、特徴量の検出と推定の処理についてのみ説明する。

## 【 0 1 1 7 】

図 4 7 のステップ  $ST_8_1$  では、画像信号  $SV$  に基づき肌色領域を抽出する。

ステップ S T 8 2 では、ステップ S T 8 1 で抽出した肌色領域の動きベクトルをブロック毎に算出する。ステップ S T 8 3 では、ステップ S T 8 2 で算出したブロック毎の動きベクトルに基づき、各ブロックの動き量をフレーム毎に平均して平均動き量を算出する。ステップ S T 8 4 では、平均動き量が、推定する観客状況に応じて設定された所定期間分算出されたか否かを判別して、所定期間分算出されていないときにはステップ S T 8 1 に戻る。また所定期間分算出されたときにはステップ S T 8 5 に進む。ステップ S T 8 5 では、所定期間分の平均動き量を平均して時間マクロ動き量を算出する。また、この時間マクロ動き量を特徴量 C W a とする。

## 【 0 1 1 8 】

図 4 8 のステップ S T 9 1 は、ステップ S T 8 3 で算出した平均動き量を取得すると共に、ステップ S T 9 2 では、平均動き量が窓幅分算出されたか否かを判別して、窓幅分算出されていないときにはステップ S T 9 1 に戻る。また窓幅分算出されたときにはステップ S T 9 3 に進む。ステップ S T 9 3 では、窓幅分の平均動き量から低周波成分比を算出する。例えば窓幅分の平均動き量のフーリエ変換やフィルタ処理を行い、全パワーに対する低周波成分のパワー比を低周波成分比として算出し、この低周波成分比を特徴量 C W b とする。

## 【 0 1 1 9 】

図 4 9 のステップ S T 1 0 1 では、フレーム期間分の音声信号 S A を抽出してステップ S T 1 0 2 に進む。ステップ S T 1 0 2 では抽出した音声信号 S A の分散を算出してステップ S T 1 0 3 に進む。ステップ S T 1 0 3 では、分散が所定期間分算出されたか否かを判別して、所定期間分算出されていないときにはステップ S T 1 0 1 に戻る。また所定期間分算出されたときにはステップ S T 1 0 4 に進む。ステップ S T 1 0 4 では、所定期間分の分散の平均を算出して、算出された時間マクロ音量を特徴量 C W c とする。

## 【 0 1 2 0 】

図 5 0 のステップ S T 1 1 1 は、ステップ S T 1 0 2 で算出した分散を取得すると共に、ステップ S T 1 1 2 では、分散が窓幅分算出されたか否かを判別して、窓幅分算出されていないときにはステップ S T 1 1 1 に戻る。また窓幅分算出

されたときにはステップ S T 1 1 3 に進む。ステップ S T 1 1 3 では、窓幅分の分散を用いて低周波成分比を算出する。例えば窓幅分の分散のフーリエ変換やフィルタ処理を行い、全パワーに対する低周波成分のパワー比を低周波成分比として算出し、この低周波成分比を特徴量 C W d とする。

#### 【 0 1 2 1 】

図 5 1 のステップ S T 1 2 1 は、窓幅分の音声信号 S A についてフーリエ変換を行う。ステップ S T 1 2 2 では、フーリエ変換結果に基づき高周波成分と低周波成分を分離してステップ S T 1 2 3 に進む。ステップ S T 1 2 3 では、高周波対低周波成分比を算出して特徴量 C W e とする。

#### 【 0 1 2 2 】

このように、動き量や動きの低周波成分、音声の音量や低周波成分、音声の高周波対低周波成分比を特徴量として検出する。

#### 【 0 1 2 3 】

次に、検出した複数の特徴量に基づく、観客状況推定処理を図 5 2 および図 5 3 に示す。なお、図 5 2 は特徴量 C W e を用いない場合、図 5 3 は特徴量 C W e を用いた場合の処理を示している。

#### 【 0 1 2 4 】

図 5 2 のステップ S T 1 3 1 では、動き量を示す特徴量 C W a が基準レベル L v 11 よりも小さいか否かを判別する。ここで、特徴量 C W a が基準レベル L v 11 よりも小さくないときにはステップ S T 1 3 2 に進む。また、基準レベル L v 11 よりも小さいときにはステップ S T 1 3 3 に進み、「その他の動作」の状態と推定する。

#### 【 0 1 2 5 】

ステップ S T 1 3 2 では、特徴量 C W a が基準レベル L v 12 よりも大きいかな否かを判別する。ここで、特徴量 C W a が基準レベル L v 12 よりも大きくないときにはステップ S T 1 3 4 に進む。また、基準レベル L v 12 よりも大きいときにはステップ S T 1 3 5 に進み、「笑い」の状態と推定する。

#### 【 0 1 2 6 】

ステップ S T 1 3 4 では、動きの周期性を示す特徴量 C W b が基準レベル L v 13

よりも大きいかな否かを判別する。ここで、特徴量  $CWb$  が基準レベル  $Lv13$  よりも大きいときにはステップ  $ST136$  に進み、「手拍子」の状態と推定する。また、基準レベル  $Lv13$  よりも大きくないときにはステップ  $ST137$  に進み、「拍手」の状態と推定する。

## 【 0 1 2 7 】

なお、図 5 2 では、動きに基づいて観客状況を推定したが、音声に基づいた特徴量  $CWc$ ,  $CWd$  を用いても同様に処理することで観客状況を推定できる。

## 【 0 1 2 8 】

高周波対低周波成分比を用いる場合、図 5 3 のステップ  $ST141$  では、動き量を示す特徴量  $CWa$  が基準レベル  $Lv11$  よりも小さいかな否かを判別する。ここで、特徴量  $CWa$  が基準レベル  $Lv11$  よりも小さくないときにはステップ  $ST142$  に進む。また、基準レベル  $Lv11$  よりも小さいときにはステップ  $ST143$  に進み、「その他の動作」の状態と推定する。

## 【 0 1 2 9 】

ステップ  $ST142$  では、高周波対低周波成分比を示す特徴量  $CWe$  が基準レベル  $La14$  よりも大きいかな否かを判別する。ここで、特徴量  $CWe$  が基準レベル  $La14$  よりも大きいときにはステップ  $ST144$  に進む。また、基準レベル  $La14$  よりも大きくないときにはステップ  $ST145$  に進み、「笑い」の状態と推定する。

## 【 0 1 3 0 】

ステップ  $ST144$  では、動きの周期性を示す特徴量  $CWb$  が基準レベル  $Lv13$  よりも大きいかな否かを判別する。ここで、特徴量  $CWb$  が基準レベル  $Lv13$  よりも大きいときにはステップ  $ST146$  に進み、「手拍子」の状態と推定する。また、基準レベル  $Lv13$  よりも大きくないときにはステップ  $ST147$  に進み、「拍手」の状態と推定する。なお、音声に基づいた特徴量  $CWc$ ,  $CWd$  を用いても同様に処理することでも観客状況を推定できる。

## 【 0 1 3 1 】

このように、動きや音声の周波数成分に基づいて特徴量を検出して、この特徴量を用いても観客状況の推定を行うことができる。さらに、動きや音声の周波数

成分に基づいて特徴量を検出する場合でも、画像信号に基づいた特徴量の推定結果と音声信号に基づいた特徴量の推定結果について論理積あるいは論理和を算出することで、推定結果が異なる場合が生じても1つの推定結果を得ることができる。また、上述の図23と同様に、視聴者60の状況の推定に用いる特徴量をコンテンツ提供状況に応じて選択して用いるものとしても良い。さらに、複数の特徴量から統合値を決定して、この統合値と閾値を比較して観客状況を推定しても良い。

【0132】

#### 【発明の効果】

この発明によれば、観客を撮影した画像信号および／または観客からの音声に応じた音声信号が生成されて、画像信号に基づき、観客の動き量や動きの周期性が検出される。また音声信号に基づき、観客からの音声の音量や音声の周期性や音声の周波数成分に関する情報が検出されて、観客状況が、検出結果に基づき推定される。このため、例えば人が観客を観察して、どのような状況であるかを判別しなくとも簡単に観客状況を推定できる。

#### 【図面の簡単な説明】

【図1】

観客状況推定装置の構成を示す図である。

【図2】

入力部を説明するための図である。

【図3】

視聴者側の撮影画像示す図である。

【図4】

特徴量検出部の構成を示す図である。

【図5】

動き量検出部の構成を示す図である。

【図6】

特徴量C V aの推移を示す図である。

【図7】

動き周期性検出部の構成を示す図である。

【図 8】

自己相関係数の算出動作を説明するための図である。

【図 9】

自己相関係数の推移を示す図である。

【図 1 0】

自己相関最大位置の推移を示す図である。

【図 1 1】

特徴量 C V b の推移を示す図である。

【図 1 2】

パワースペクトル検出部の構成を示す図である。

【図 1 3】

音声信号を示す図である。

【図 1 4】

音声信号のフーリエ変換結果を示す図である。

【図 1 5】

特徴量 C V c の推移を示す図である。

【図 1 6】

特徴量 C V d の推移を示す図である。

【図 1 7】

音声周期性検出部の構成を示す図である。

【図 1 8】

自己相関最大位置の推移を示す図である。

【図 1 9】

特徴量 C V e の推移を示す図である。

【図 2 0】

推定動作を示す図である。

【図 2 1】

特徴量に基づいた推定結果（その 1）を示す図である。

【図 2 2】

特徴量に基づいた推定結果（その 2）を示す図である。

【図 2 3】

観客状況の推定に用いる特徴量とコンテンツ提示状況の関係を示す図である。

【図 2 4】

コンピュータを用いた構成を示す図である。

【図 2 5】

観客状況推定処理の全体を示すフローチャートである。

【図 2 6】

特徴量 C V a の検出処理を示すフローチャートである。

【図 2 7】

特徴量 C V b の検出処理を示すフローチャートである。

【図 2 8】

特徴量 C V c の検出処理を示すフローチャートである。

【図 2 9】

特徴量 C V d の検出処理を示すフローチャートである。

【図 3 0】

特徴量 C V e の検出処理を示すフローチャートである。

【図 3 1】

観客状況の推定処理を示すフローチャートである。

【図 3 2】

特徴量検出部の他の構成を示す図である。

【図 3 3】

動き量検出部の構成を示す図である。

【図 3 4】

特徴量 C W a の推移を示す図である。

【図 3 5】

平均動き量のフーリエ変換結果を示す図である。

【図 3 6】



特徴量C Wbの推移を示す図である。

【図 3 7】

音量検出部の構成を示す図である。

【図 3 8】

特徴量C Wcの推移を示す図である。

【図 3 9】

特徴量C Wdの推移を示す図である。

【図 4 0】

特徴量C Weの推移を示す図である。

【図 4 1】

推定処理（特徴量C Weを用いない場合）を示す図である。

【図 4 2】

特徴量に基づいた推定結果（その 3）を示す図である。

【図 4 3】

特徴量に基づいた推定結果（その 4）を示す図である。

【図 4 4】

特徴量に基づいた推定結果（その 5）を示す図である。

【図 4 5】

推定処理（特徴量C Weを用いる場合）を示す図である。

【図 4 6】

特徴量に基づいた推定結果（その 6）を示す図である。

【図 4 7】

特徴量C Waの検出処理を示すフローチャートである。

【図 4 8】

特徴量C Wbの検出処理を示すフローチャートである。

【図 4 9】

特徴量C Wcの検出処理を示すフローチャートである。

【図 5 0】

特徴量C Wdの検出処理を示すフローチャートである。

【図 5 1】

特徴量 C We の検出処理を示すフローチャートである。

【図 5 2】

観客状況の推定処理（特徴量 C We を用いない場合）を示すフローチャートである。

【図 5 3】

観客状況の推定処理（特徴量 C We を用いる場合）を示すフローチャートである。

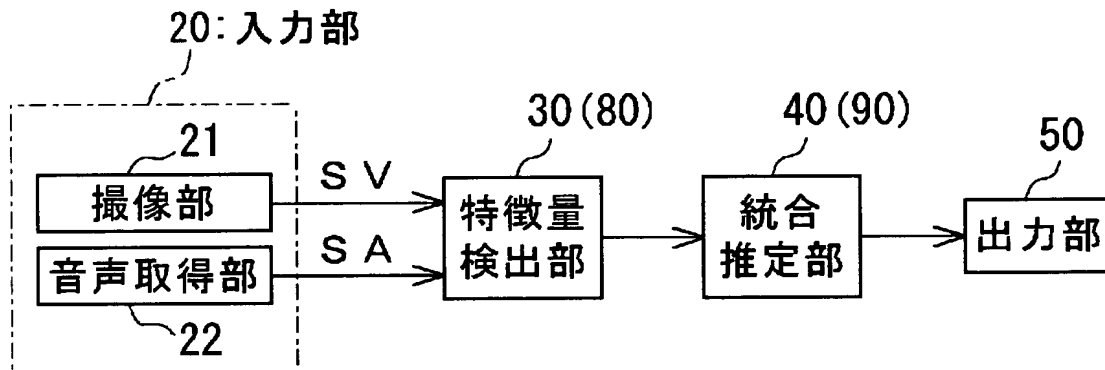
【符号の説明】

1 0 . . . 観客状況推定装置、2 0 . . . 入力部、2 1 . . . 撮像部、2 2 . . . 音声取得部、3 0, 8 0 . . . 特徴量検出部、3 1, 8 1 . . . 動き量検出部、3 2, 8 2 . . . 周期性検出部、3 3 . . . フィルタ部、3 4 . . . パワースペクトル検出部、3 5, 8 4 . . . 音量検出部、3 6, 8 5 . . . 音声周期性検出部、4 0, 9 0 . . . 推定部、5 0 . . . 出力部、6 0 . . . 視聴者、6 1 . . . 画像表示装置、6 2 . . . スピーカ、8 6 . . . 高周波対低周波成分比計算部、3 1 1 . . . 肌色領域抽出部、3 1 2 . . . 動きベクトル計算部、3 1 3 . . . 平均動き量計算部、3 2 1, 3 6 1 . . . 自己相関計算部、3 2 2, 3 6 2 . . . 最大位置抽出部、3 2 3, 3 6 3, 8 4 1 . . . 分散計算部、3 4 1 . . . フーリエ変換部、3 4 2 . . . 重心計算部、8 1 1, 8 4 2 . . . 時間平均計算部、8 5 1 . . . フーリエ変換部、8 5 2 . . . 低周波成分比計算部

【書類名】 図面

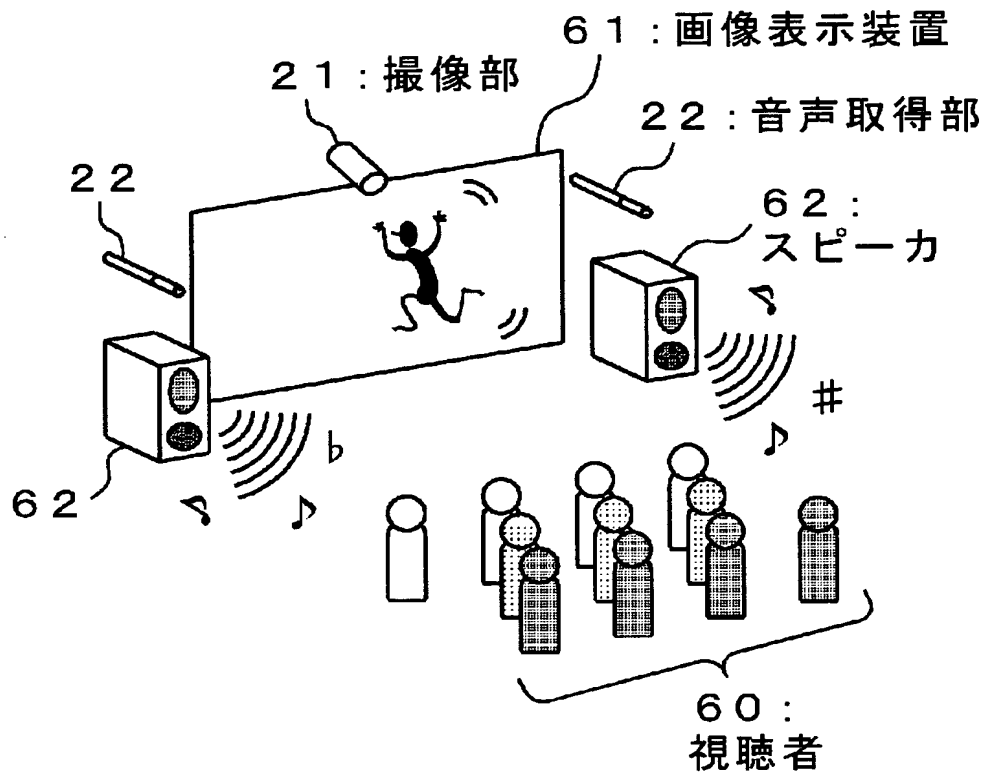
【図 1】

## 観客状況推定装置の構成



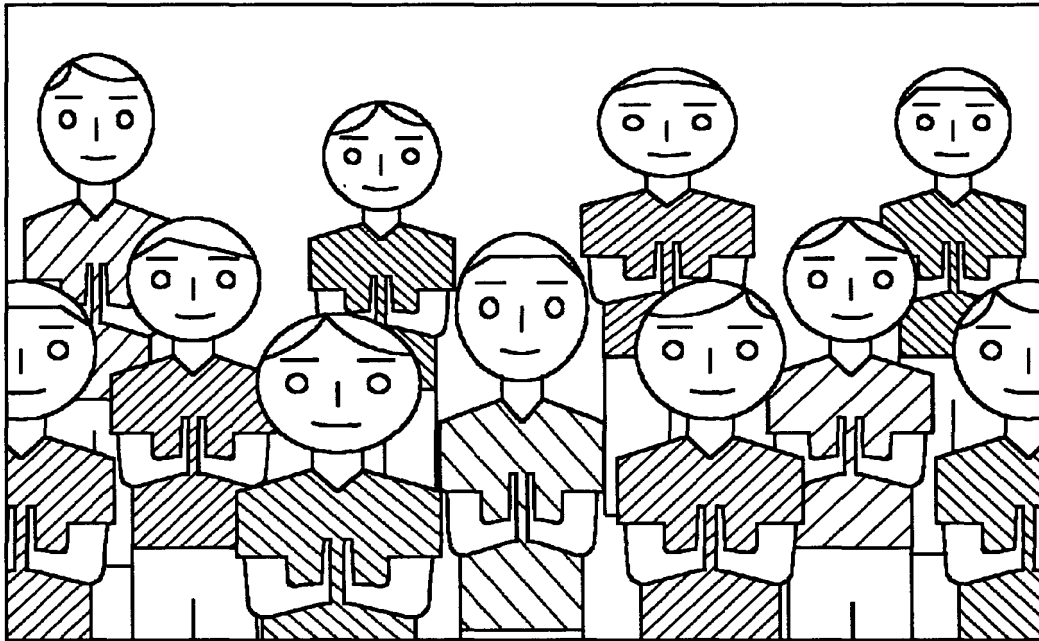
【図2】

## 入力部



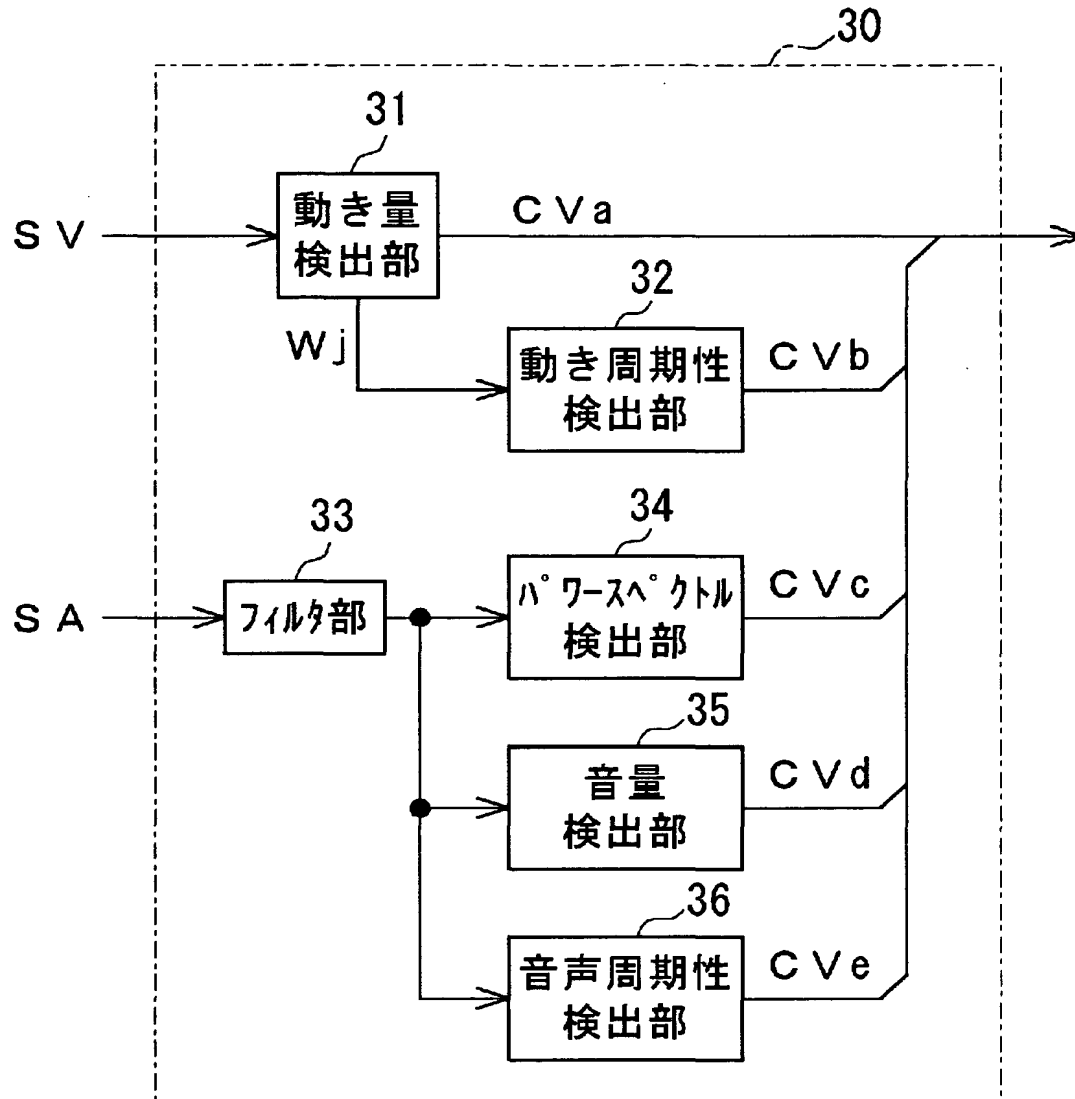
【図3】

## 視聴者側の撮影画像



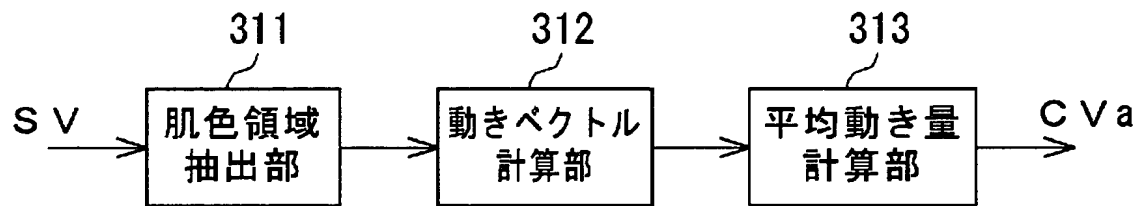
【図 4】

# 特徴量検出部の構成



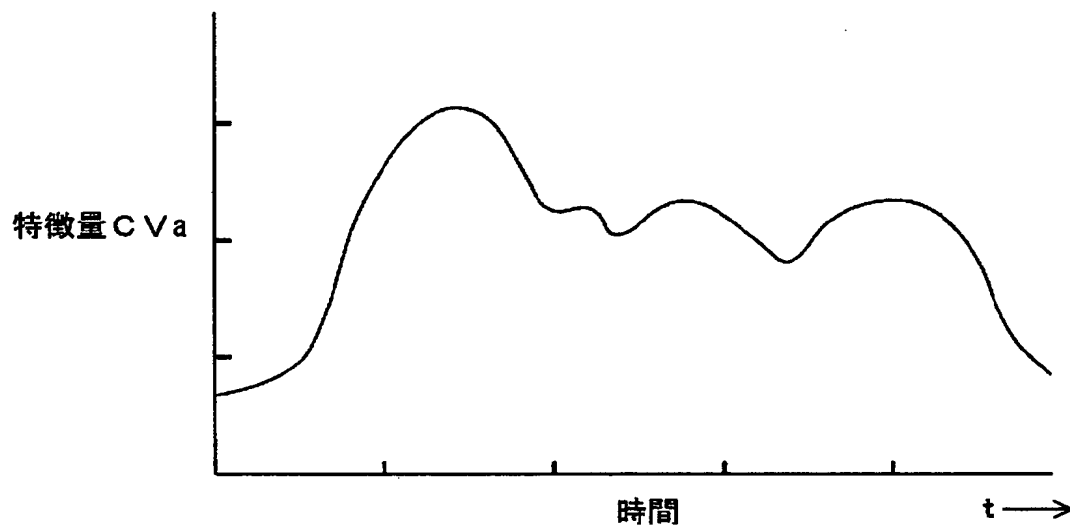
【図 5】

## 動き量検出部の構成



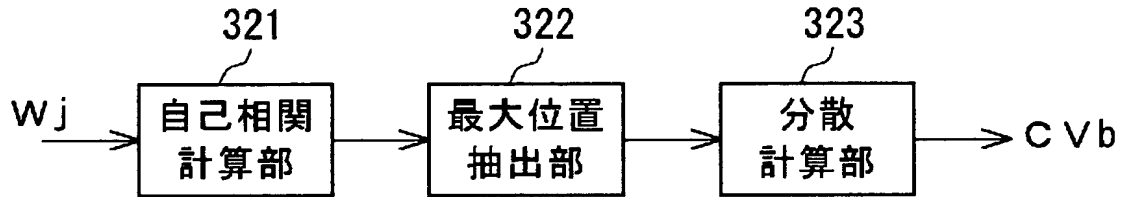
【図 6】

## 特徴量 CVa の推移



【図 7】

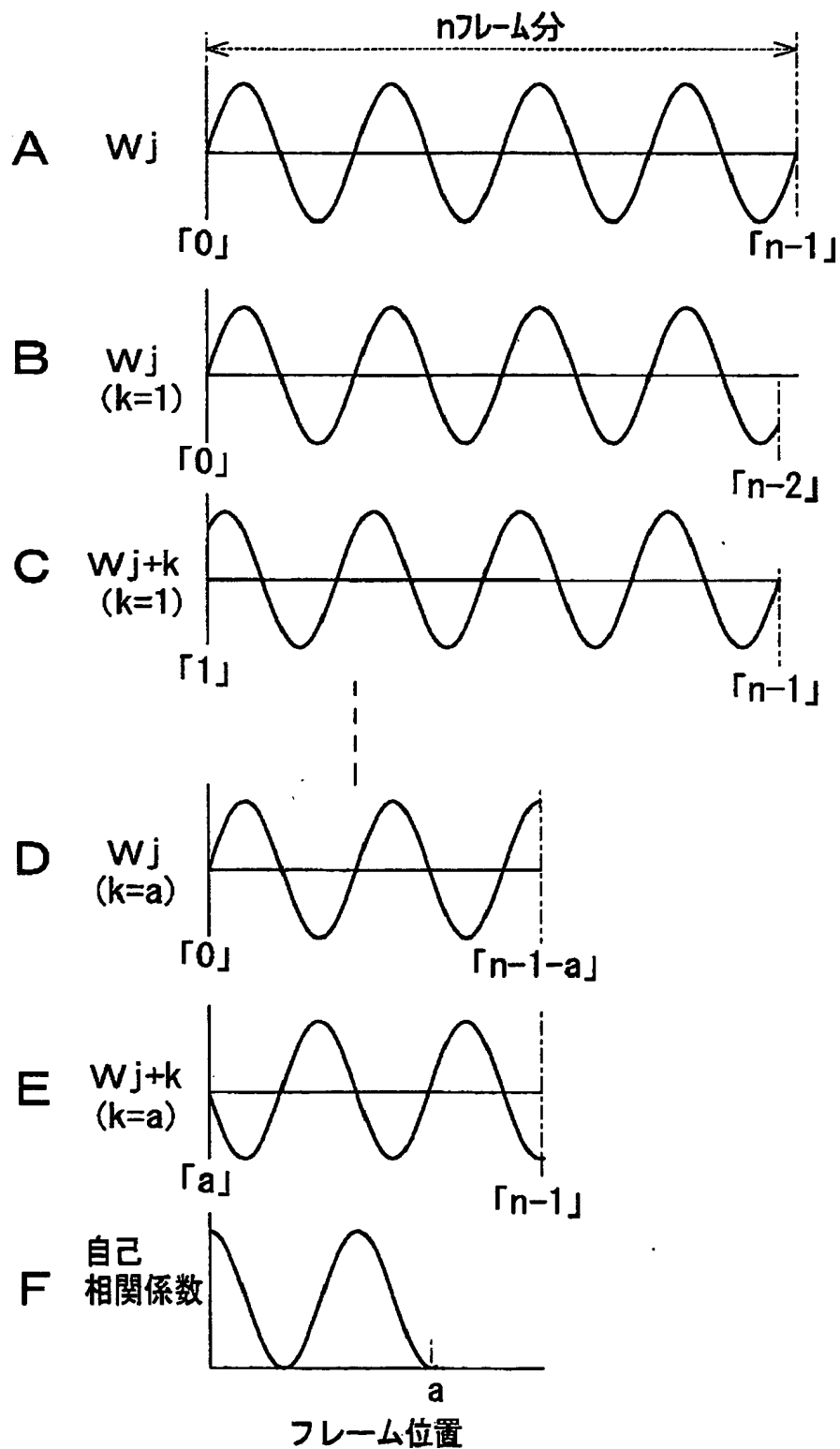
## 動き周期性検出部の構成





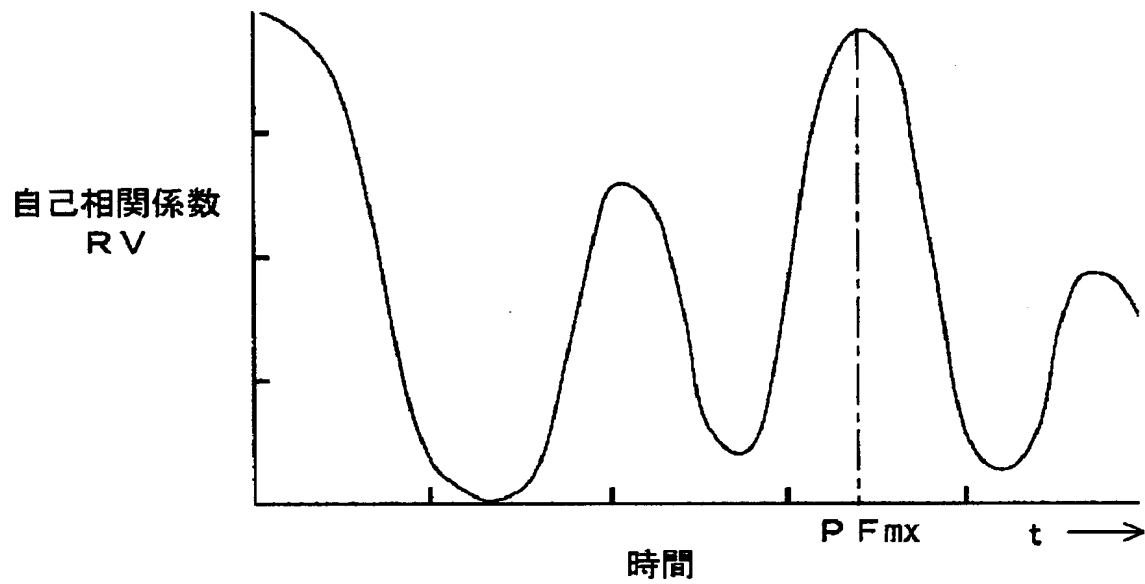
【図 8】

# 自己相関係数の算出動作



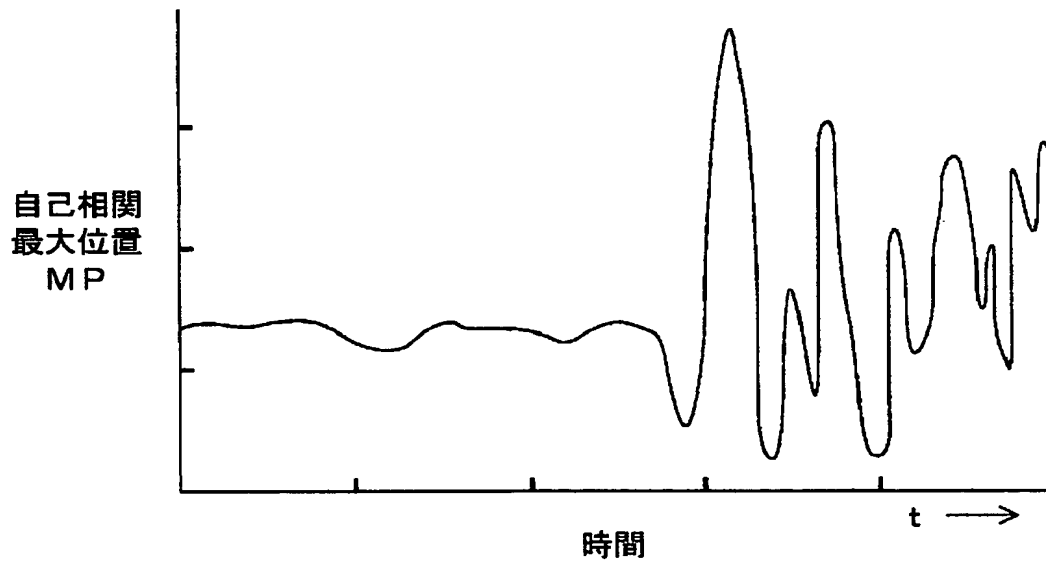
【図 9】

## 自己相関係数の推移



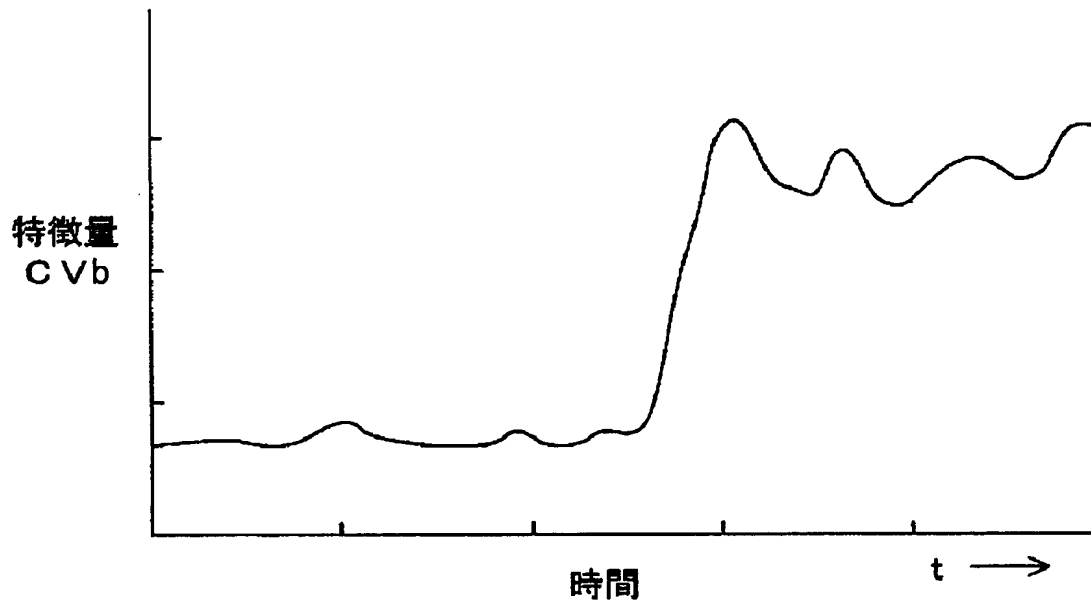
【図 1 0】

## 自己相関最大位置の推移



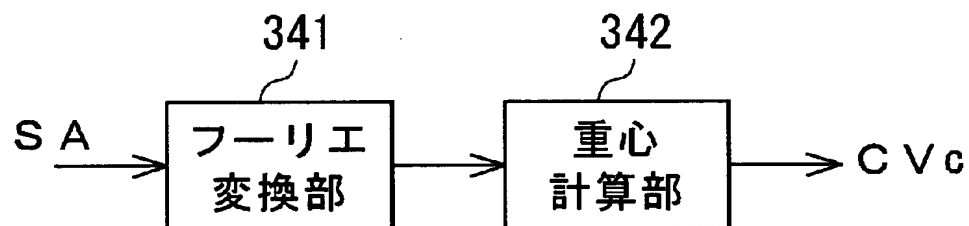
【図 1 1】

## 特徴量 C V b の推移



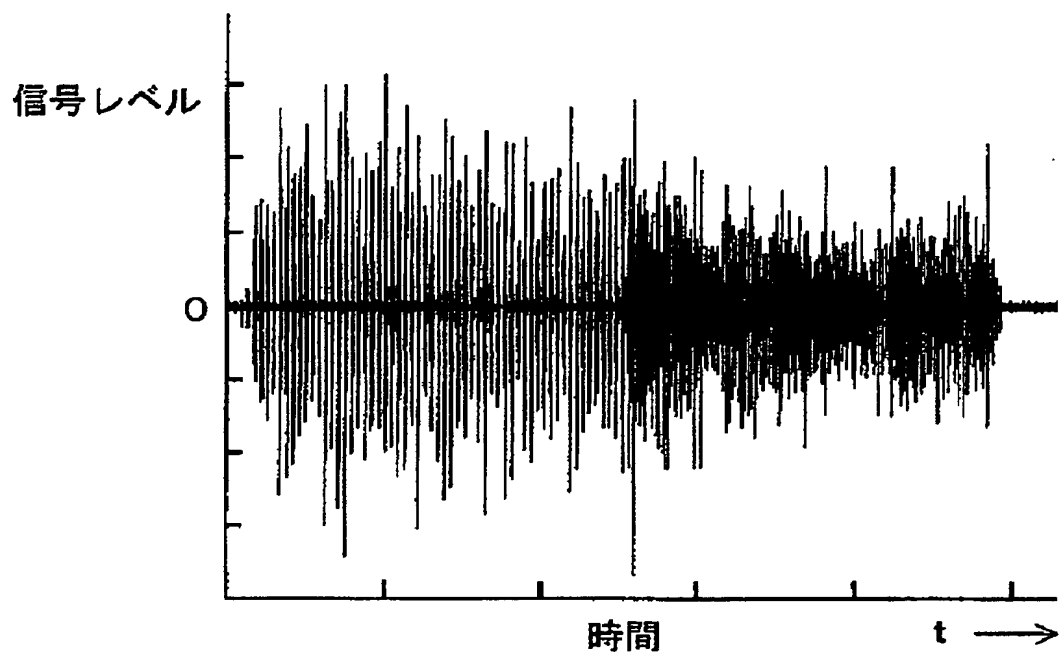
【図 1 2】

## パワースペクトル検出部の構成



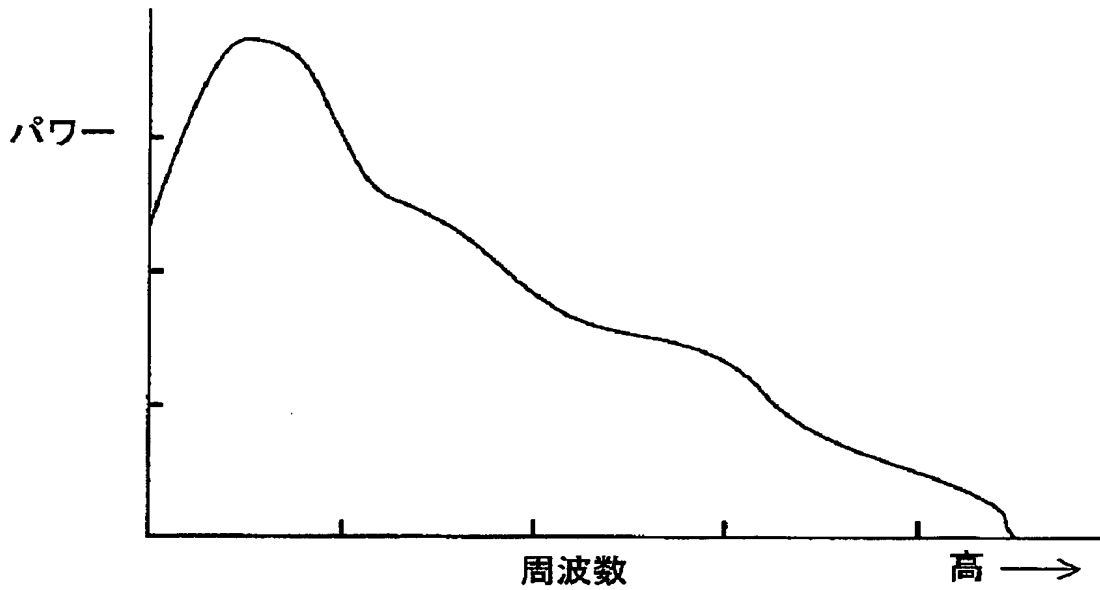
【図 1 3】

## 音声信号



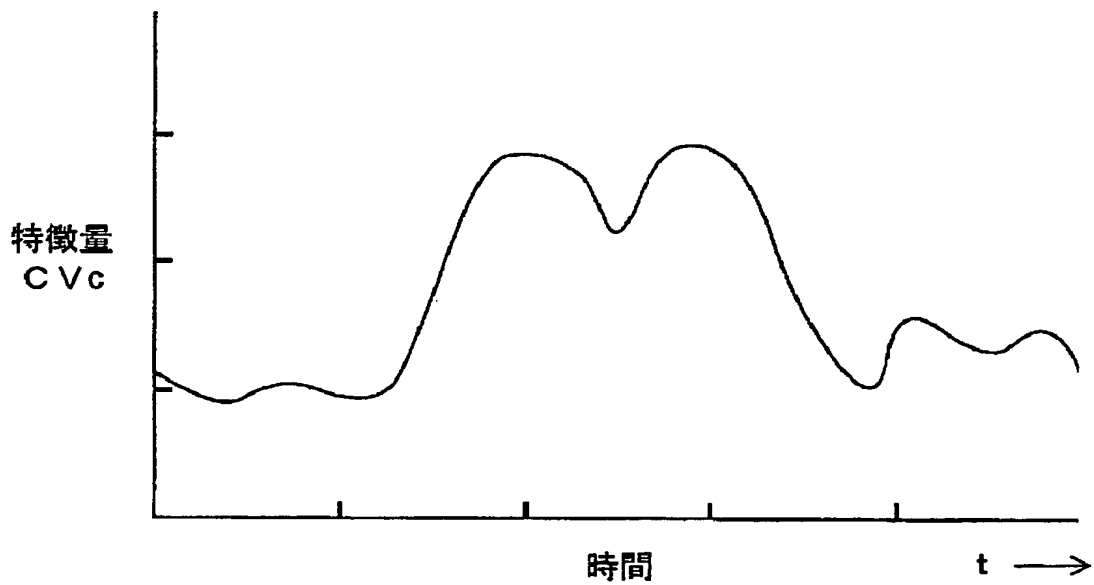
【図 1 4】

## フーリエ変換結果



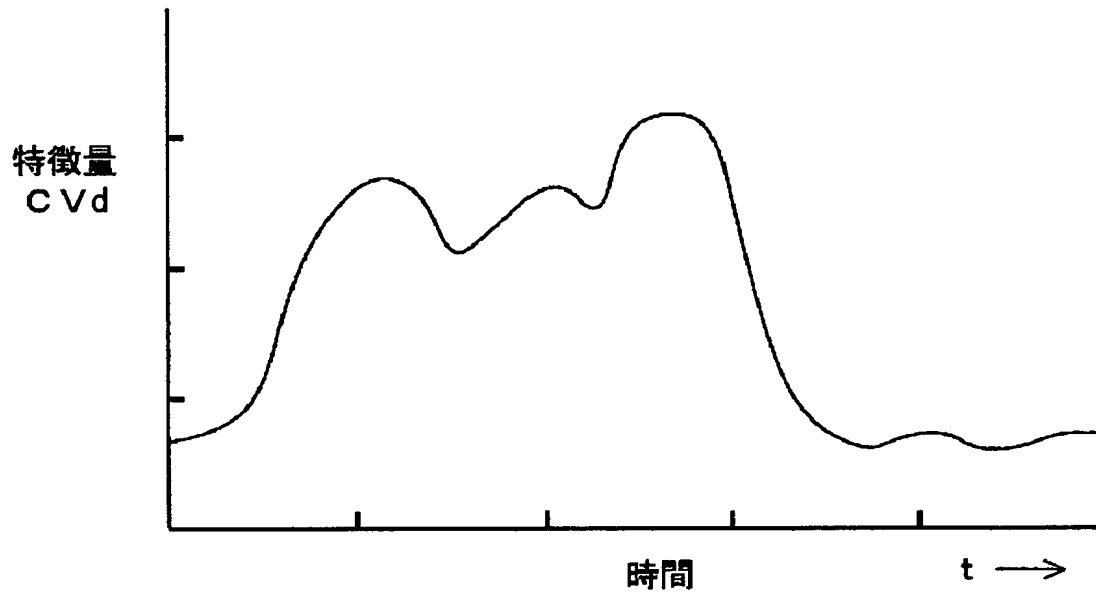
【図 1 5】

## 特徴量 $CV_c$ の推移



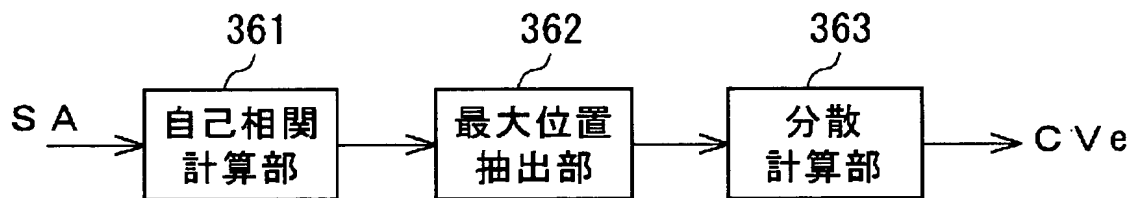
【図 1 6】

## 特徴量 C V d の推移



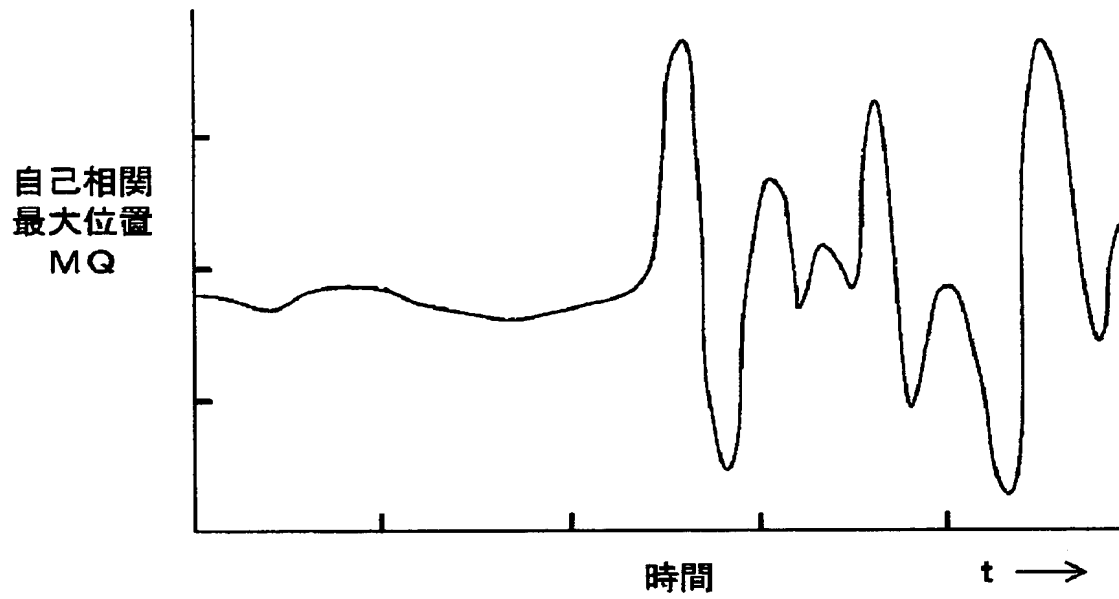
【図 1 7】

## 音声周期性検出部



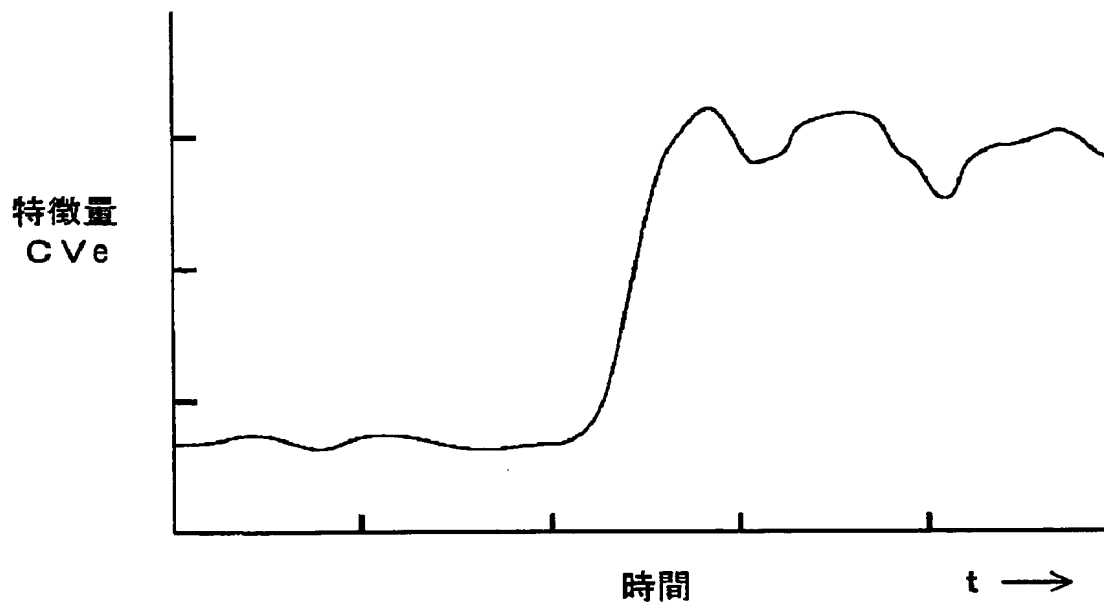
【図 1 8】

## 自己相関最大位置の推移



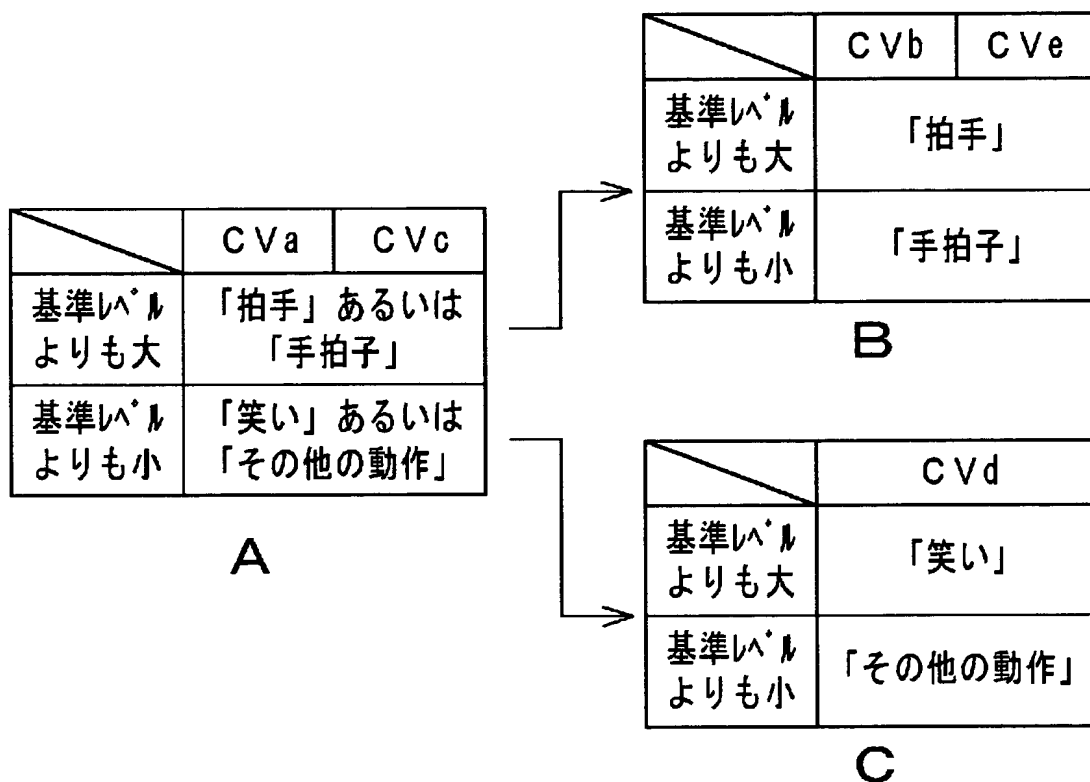
【図 1 9】

## 特徴量 C V e の推移



【図 2 0】

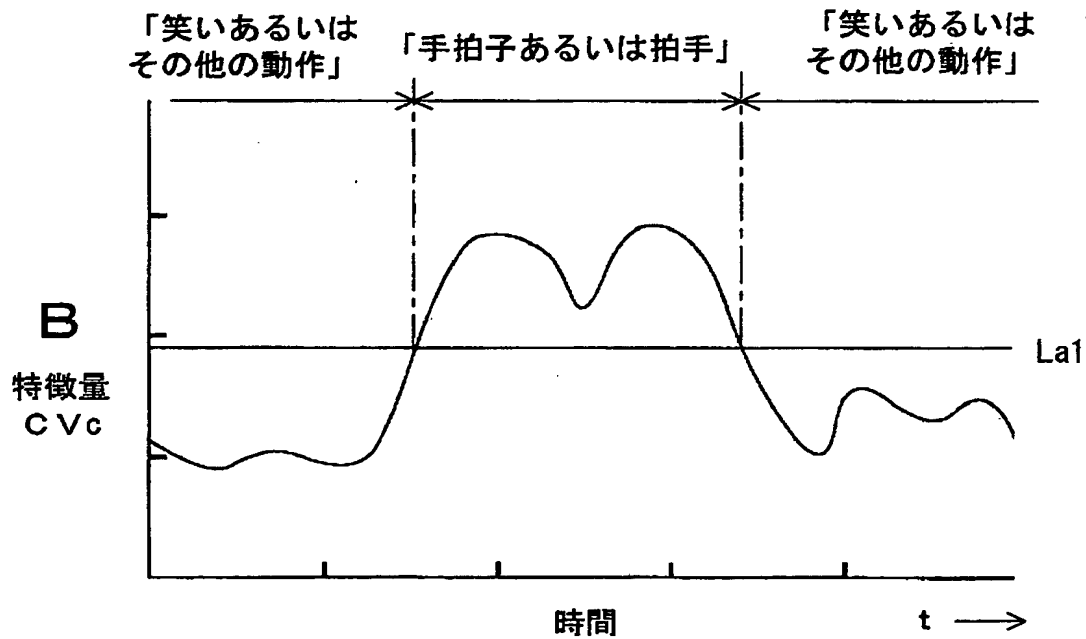
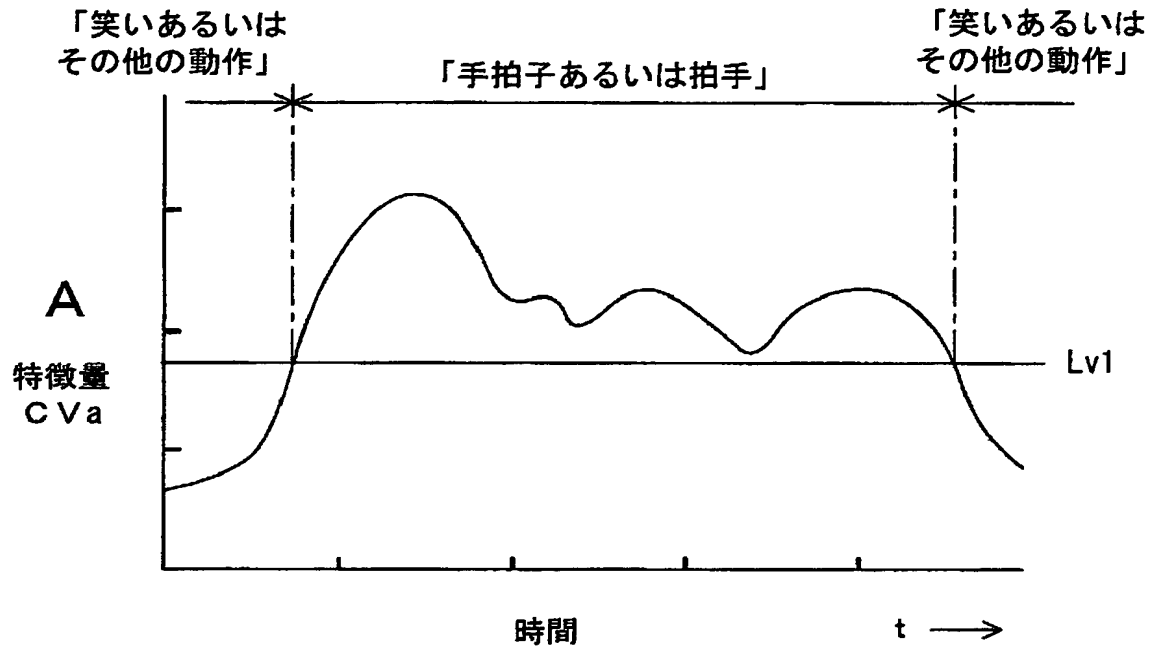
## 推定動作





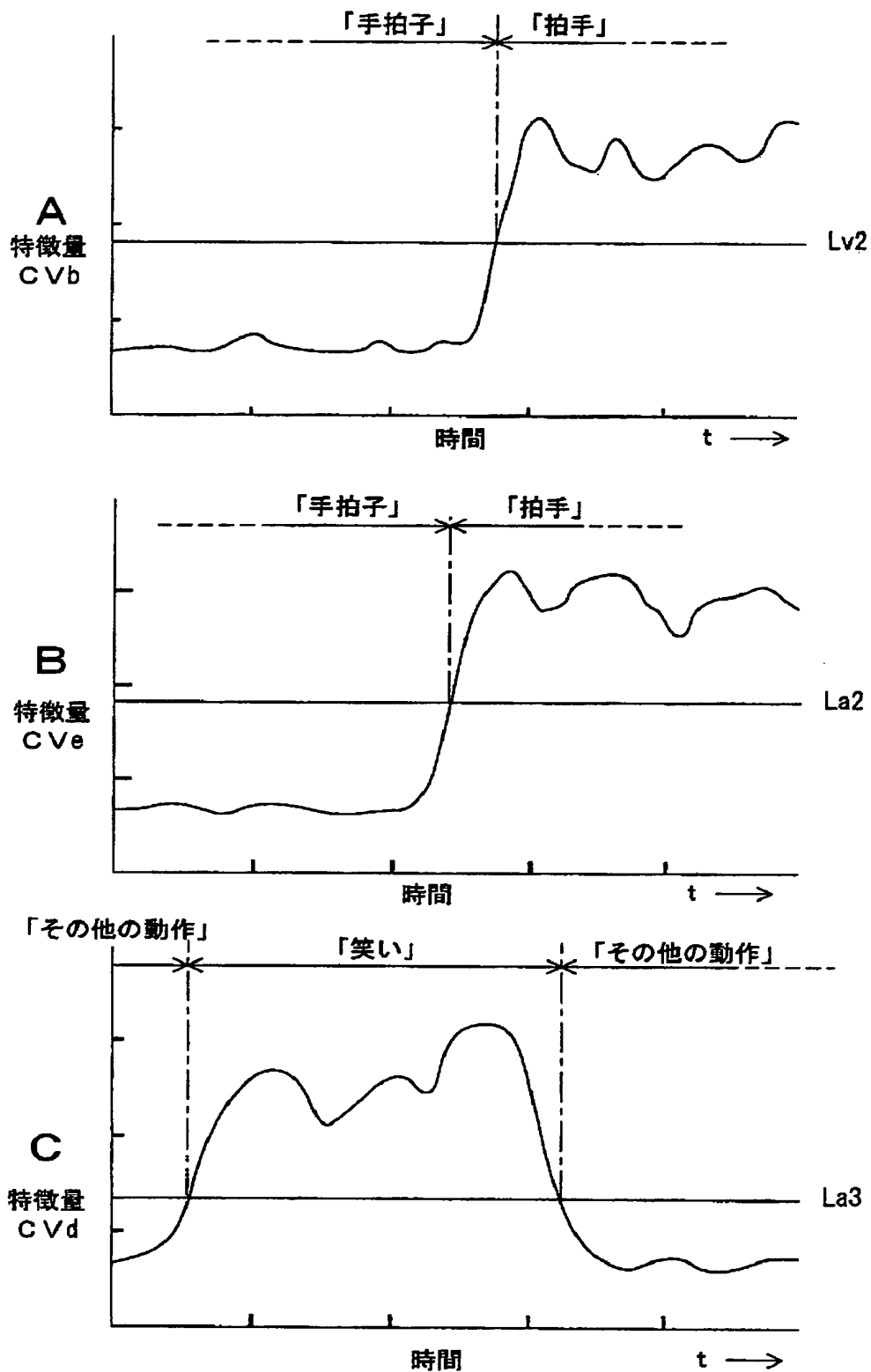
【図 21】

# 特徴量に基づいた推定結果（その 1）



【図 2 2】

特徴量に基づいた推定結果（その 2）



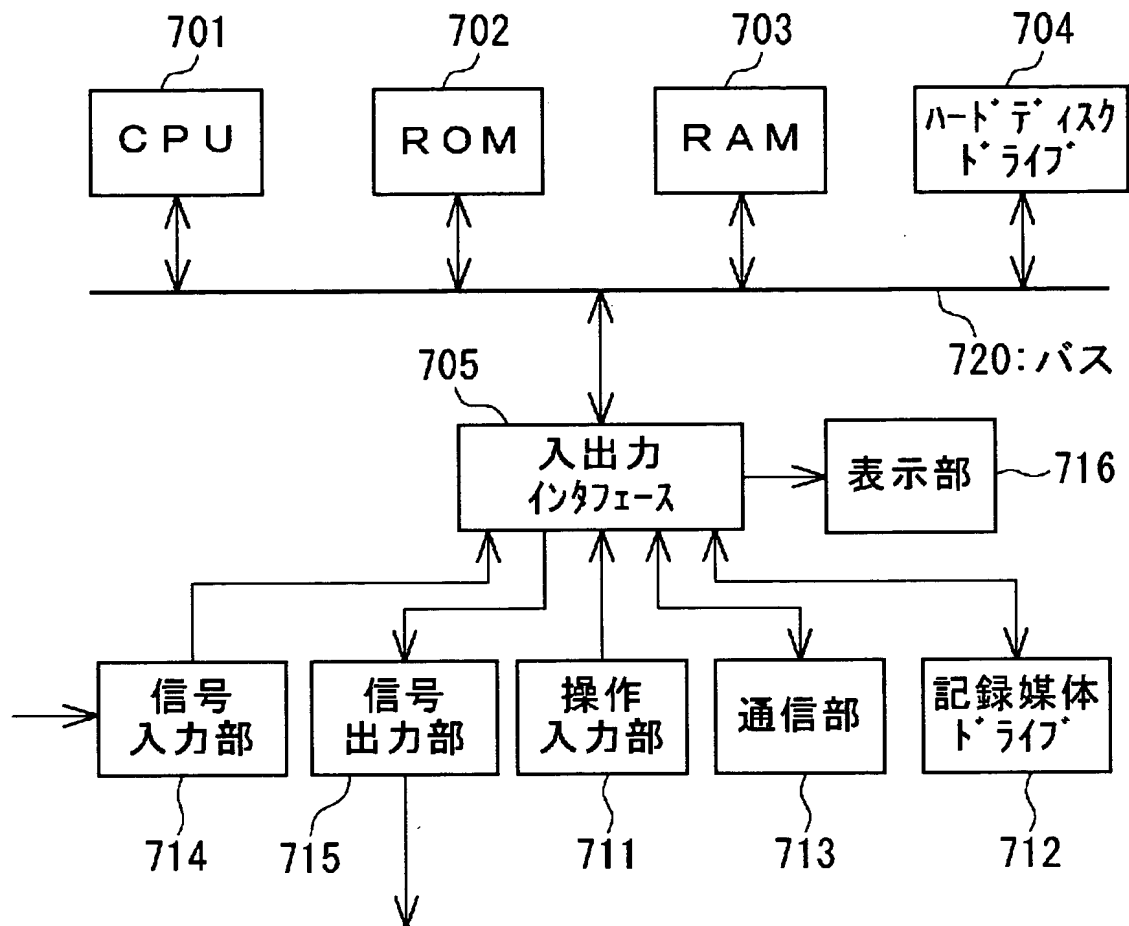
【図 2 3】

## 観客状況の推定に用いる特徴量と コンテンツの提示状況

	推定に用いる特徴量				
	手拍子・拍手と笑い ・その他の動作の 推定時		手拍子・拍手の 推定時		C V d
	C V a	C V c	C V b	C V e	
コンサート	○			○	○
映画館		○		○	○
スポーツ中継	○		○		○
⋮					

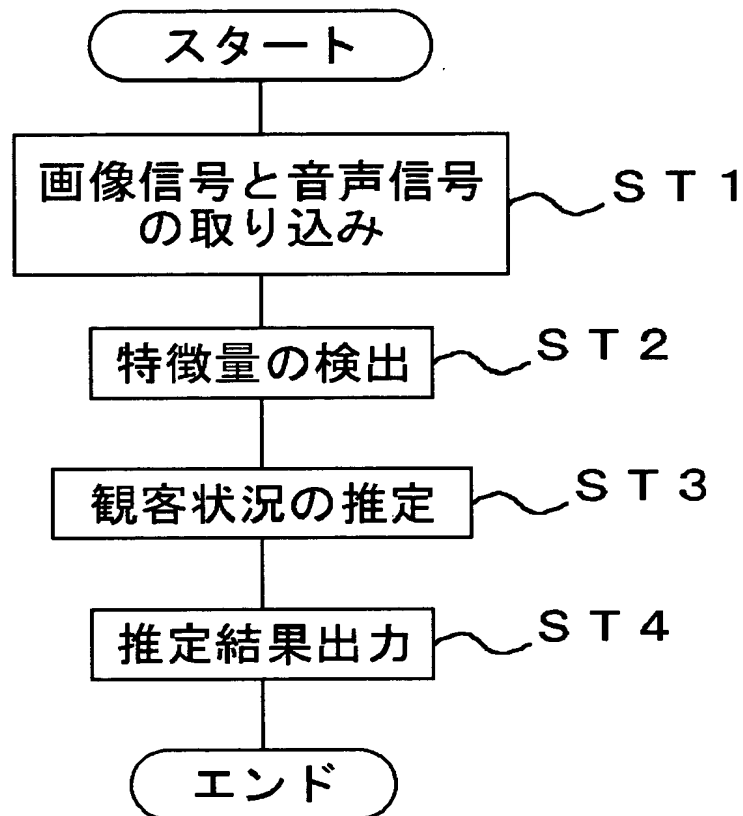
【図 2 4】

# コンピュータを用いた構成



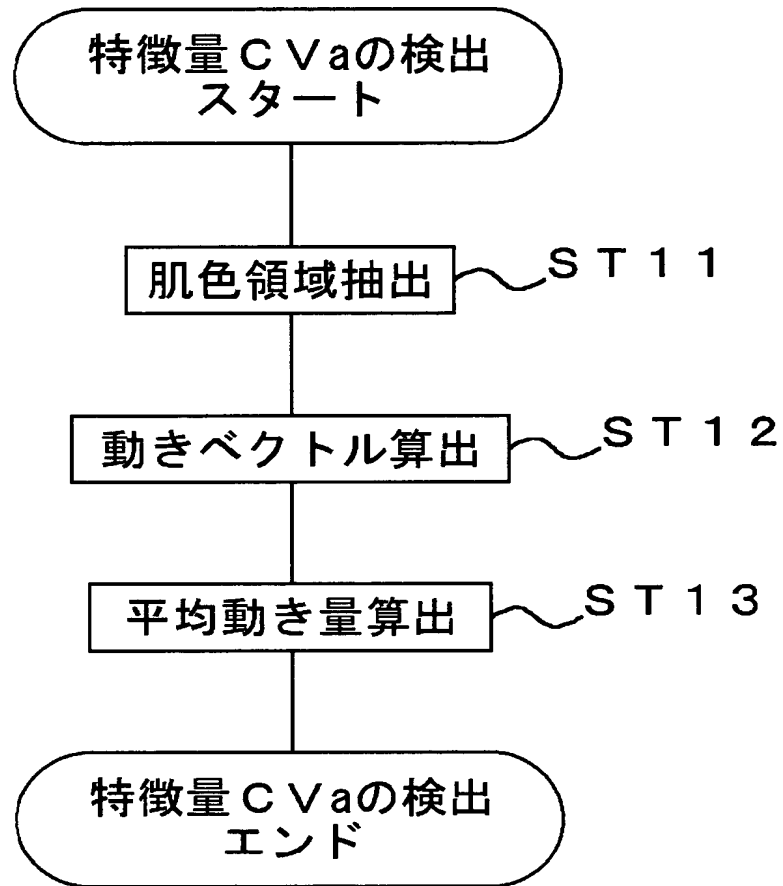
【図 2 5】

## 観客状況推定処理



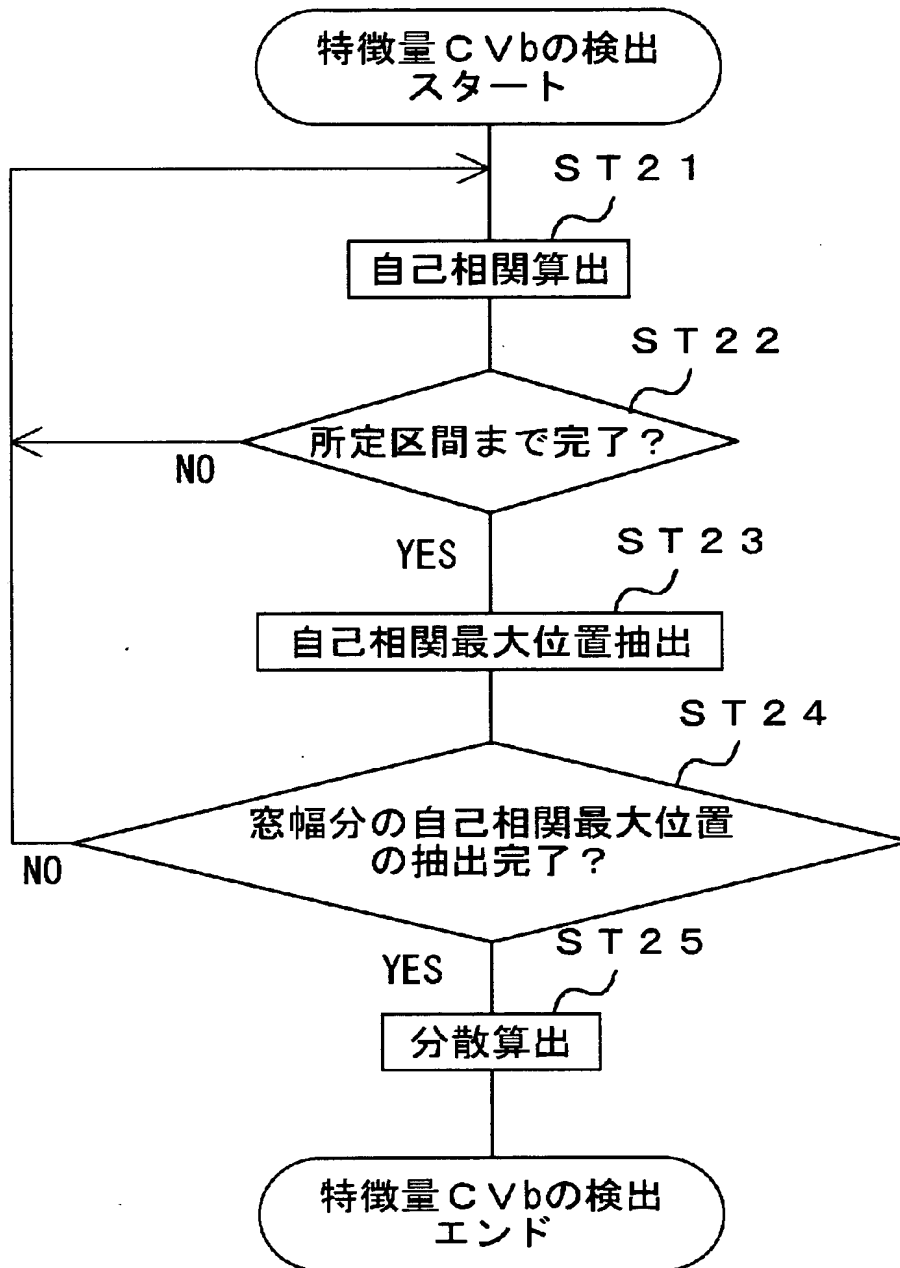
【図 2 6】

## 特徴量 C V a の検出処理



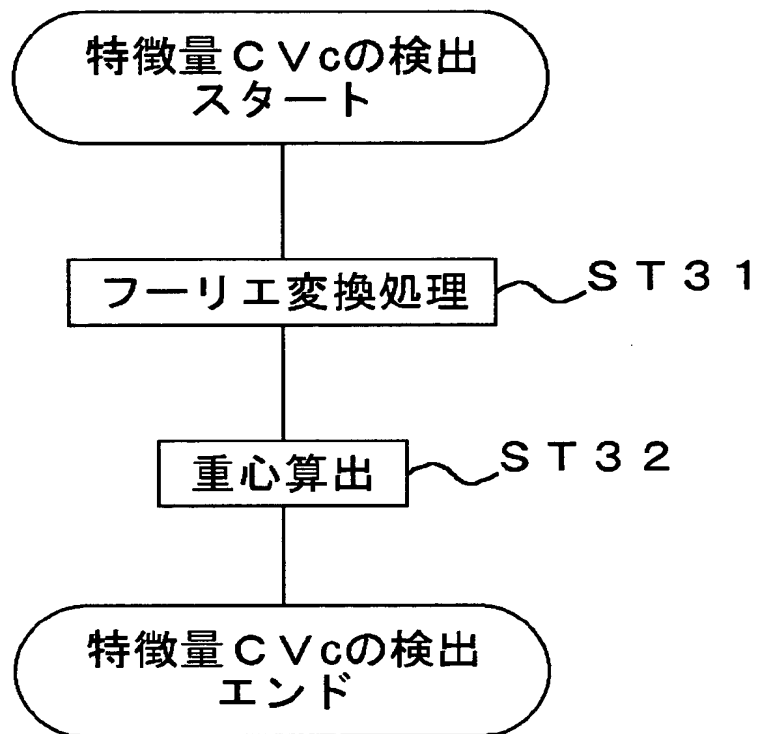
【図 2 7】

## 特徴量 C V b の検出処理



【図 2 8】

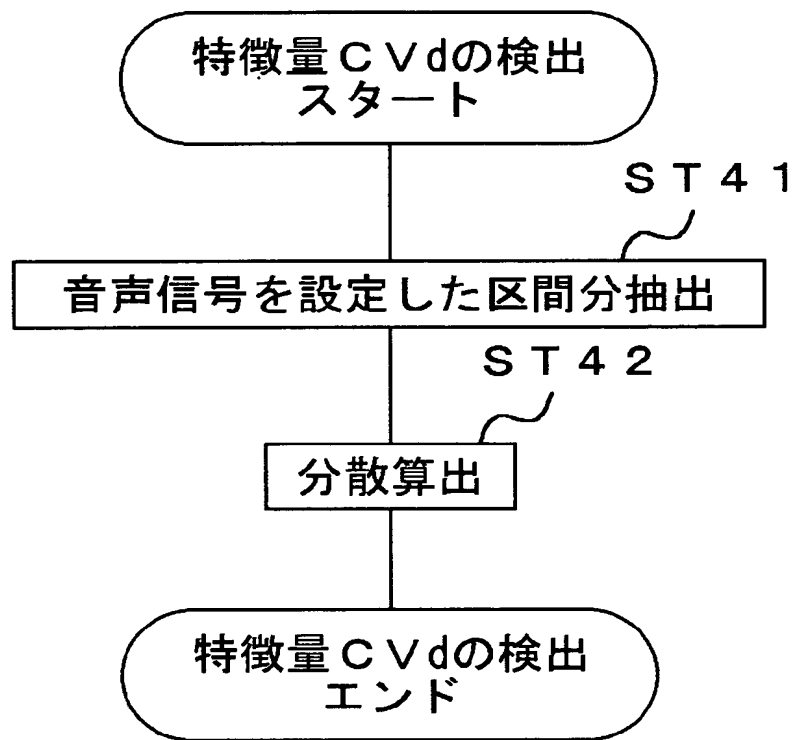
## 特徴量 C V c の検出処理





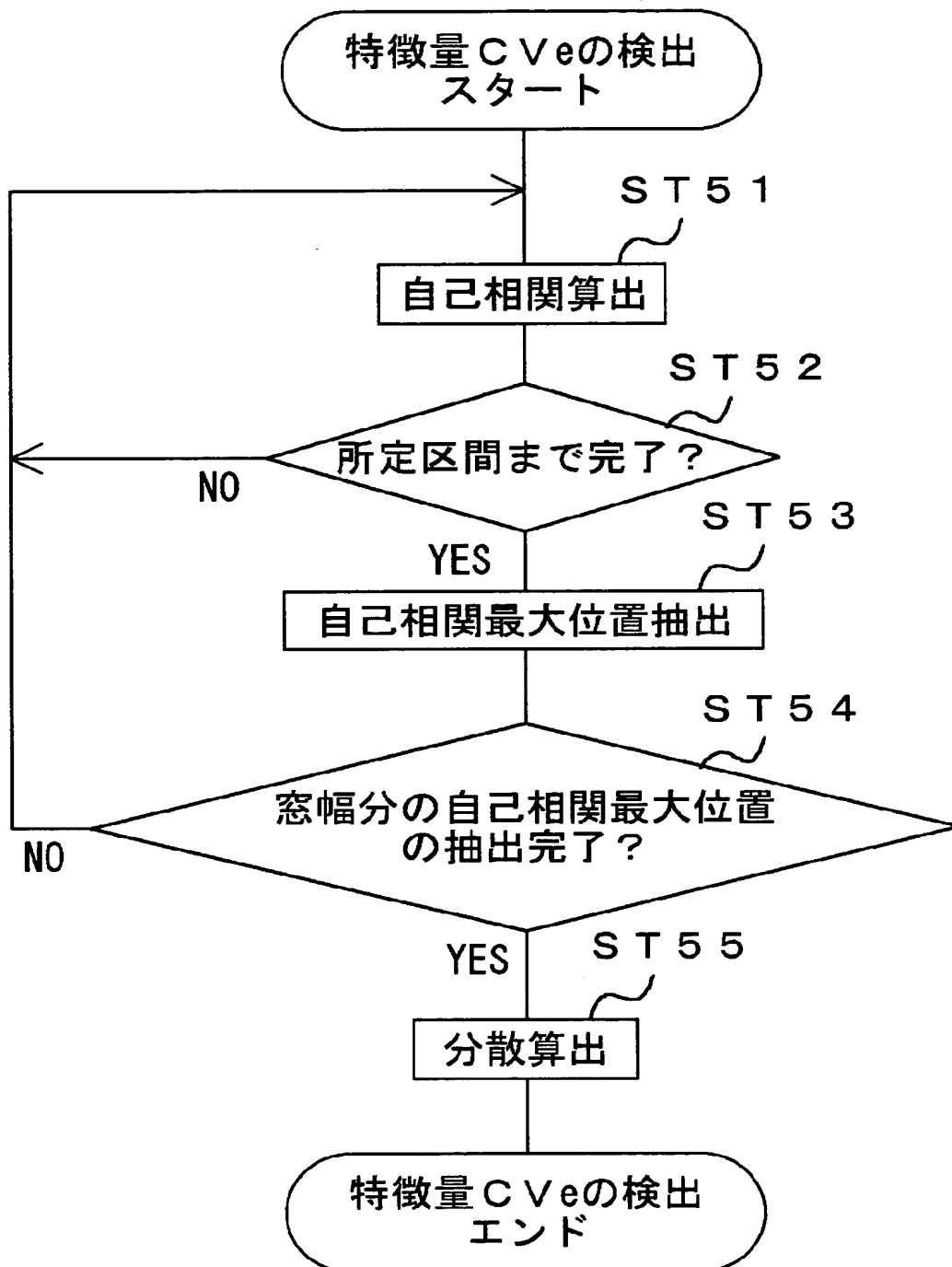
【図 2 9】

## 特徴量 C V d の検出処理



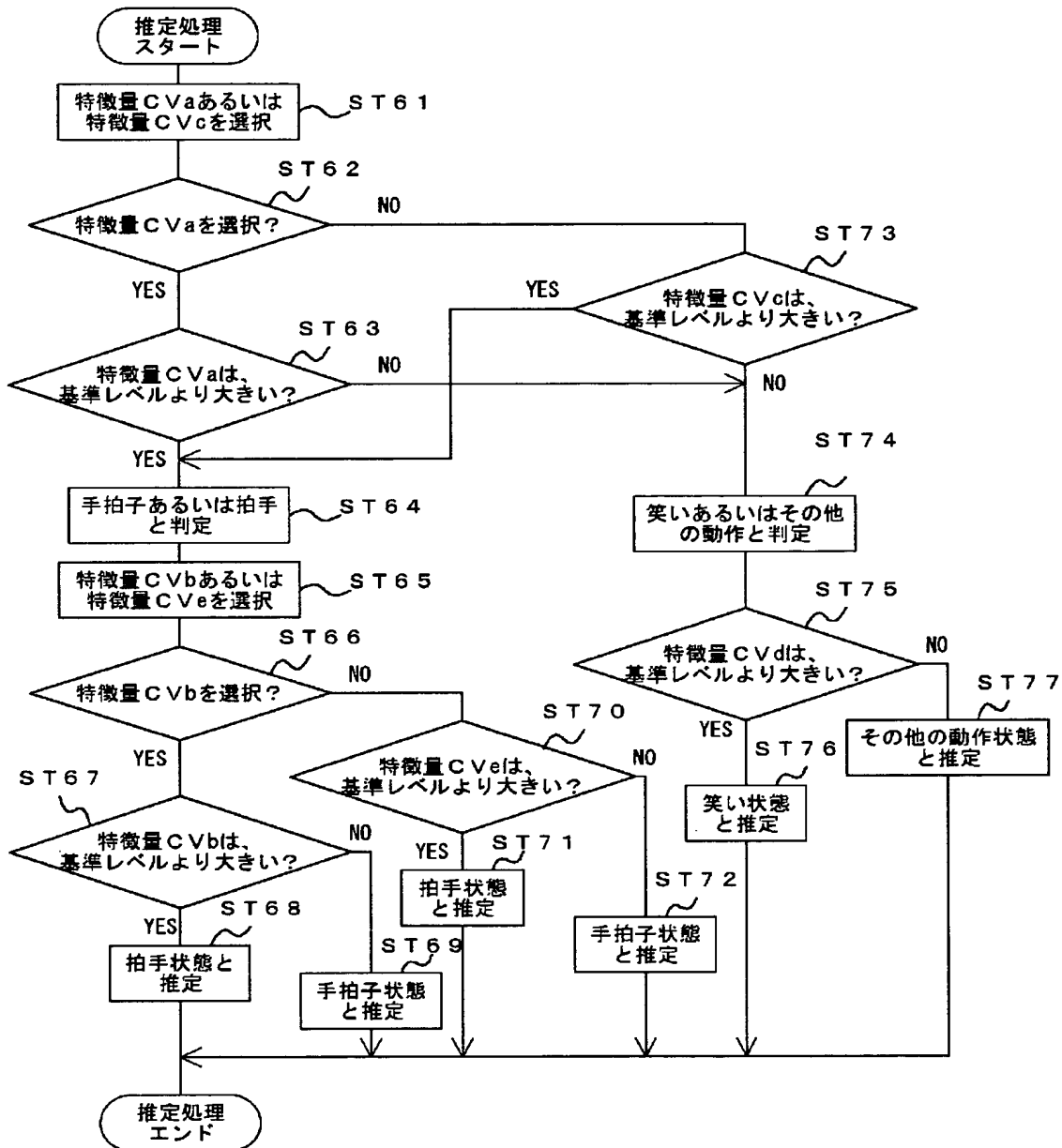
【図 3 0】

# 特徴量 C V e の検出処理



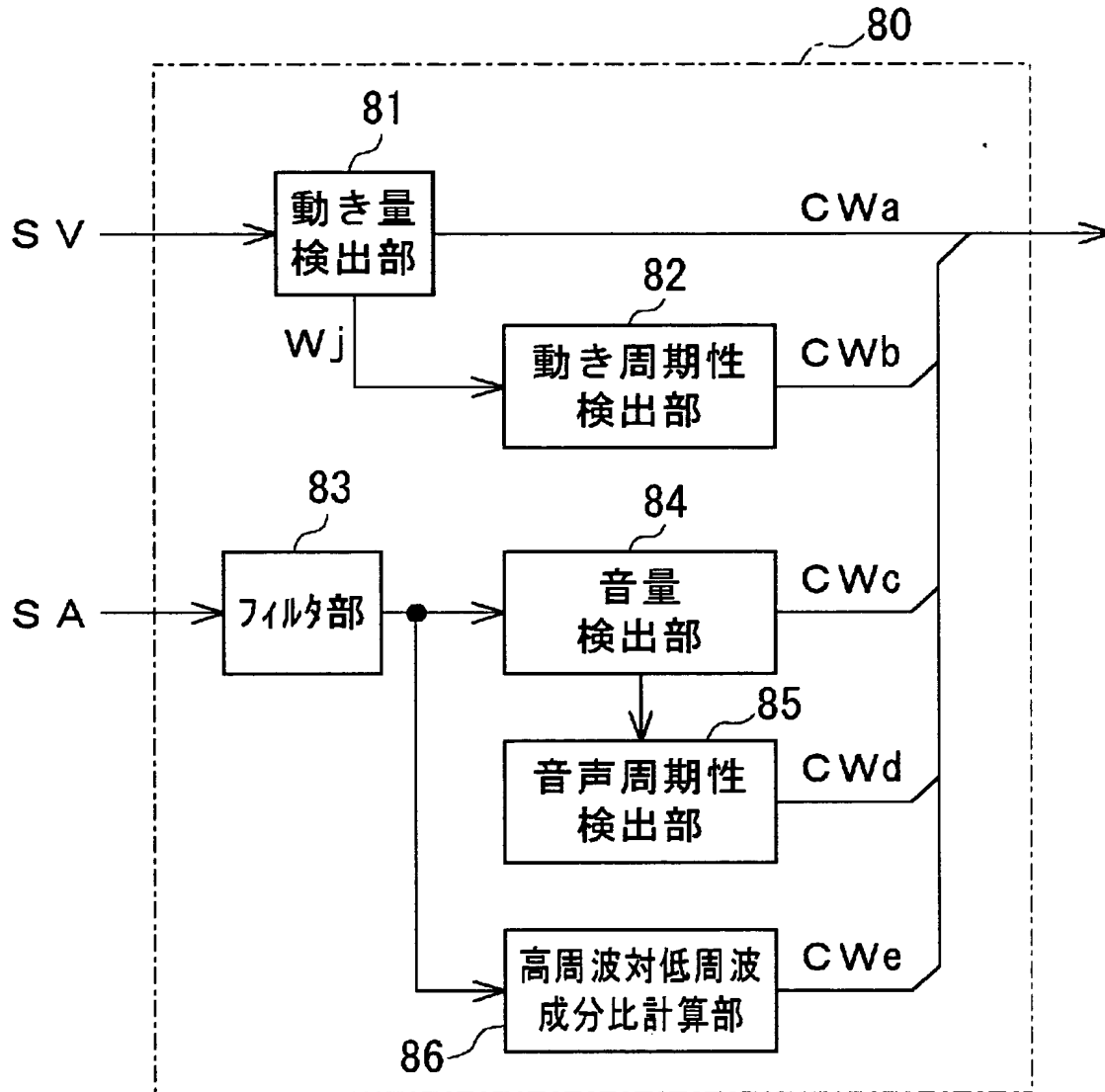
【図 31】

# 観客状況の推定処理



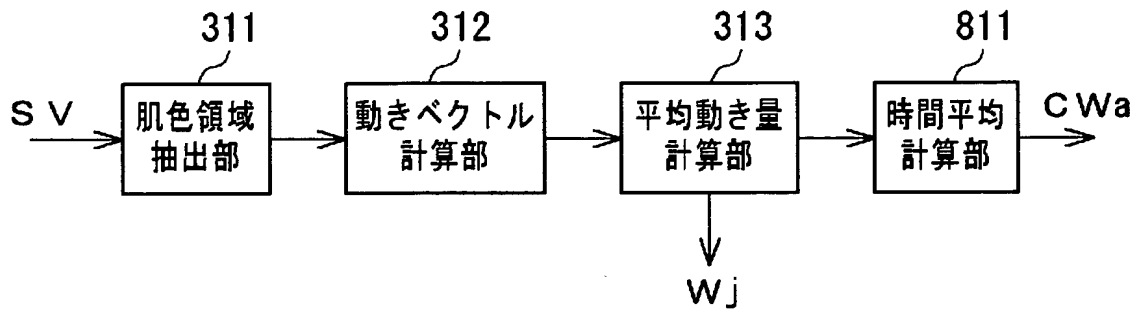
【図 3 2】

# 特徴量検出部の他の構成



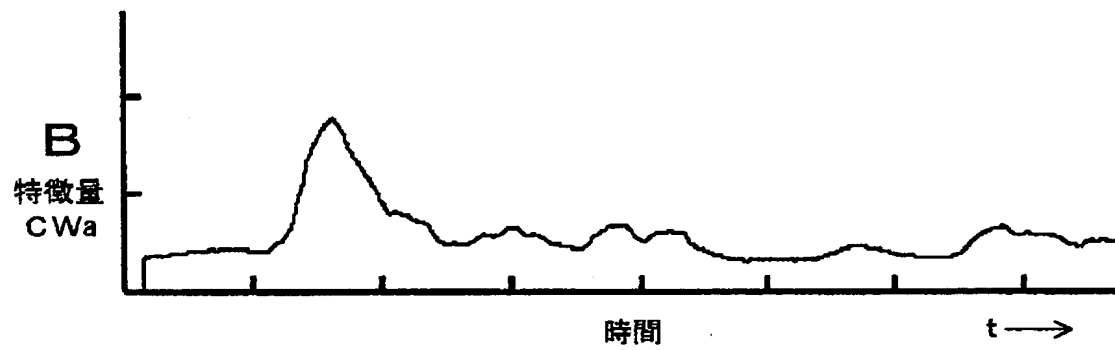
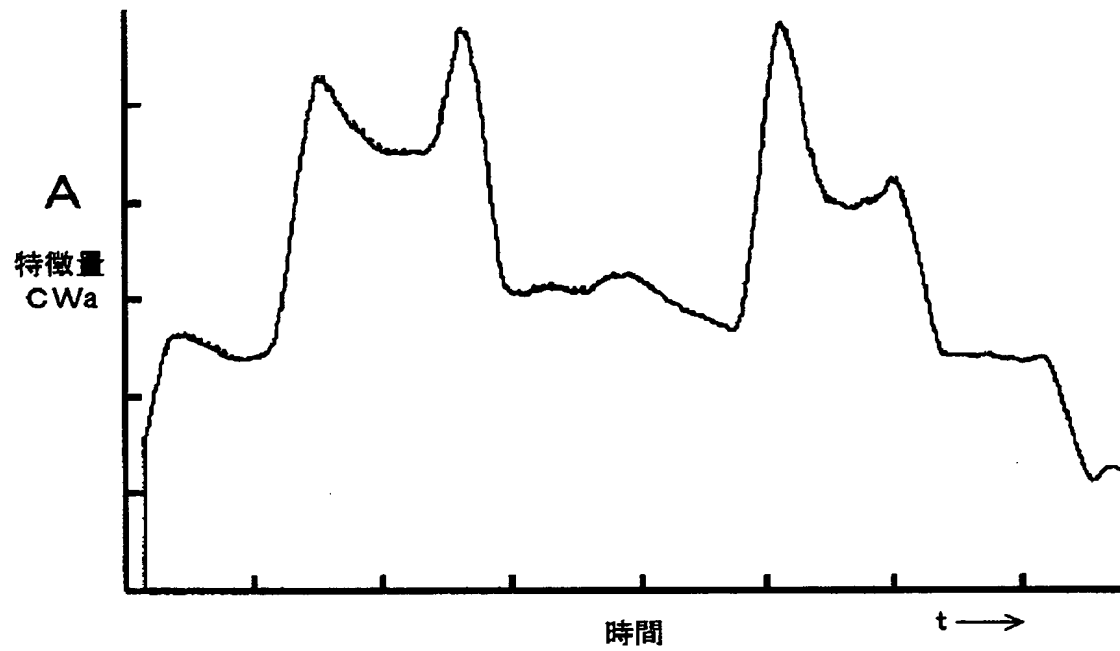
【図 3 3】

# 動き量検出部の構成



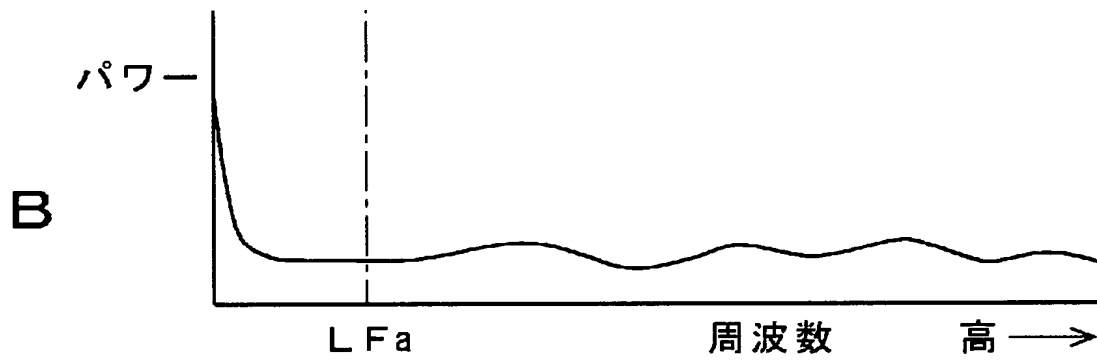
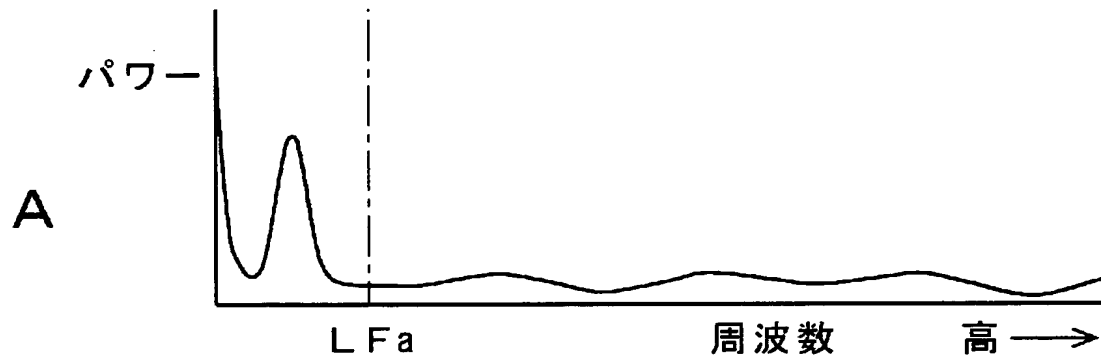
【図 3 4】

# 特徴量 C W a の推移



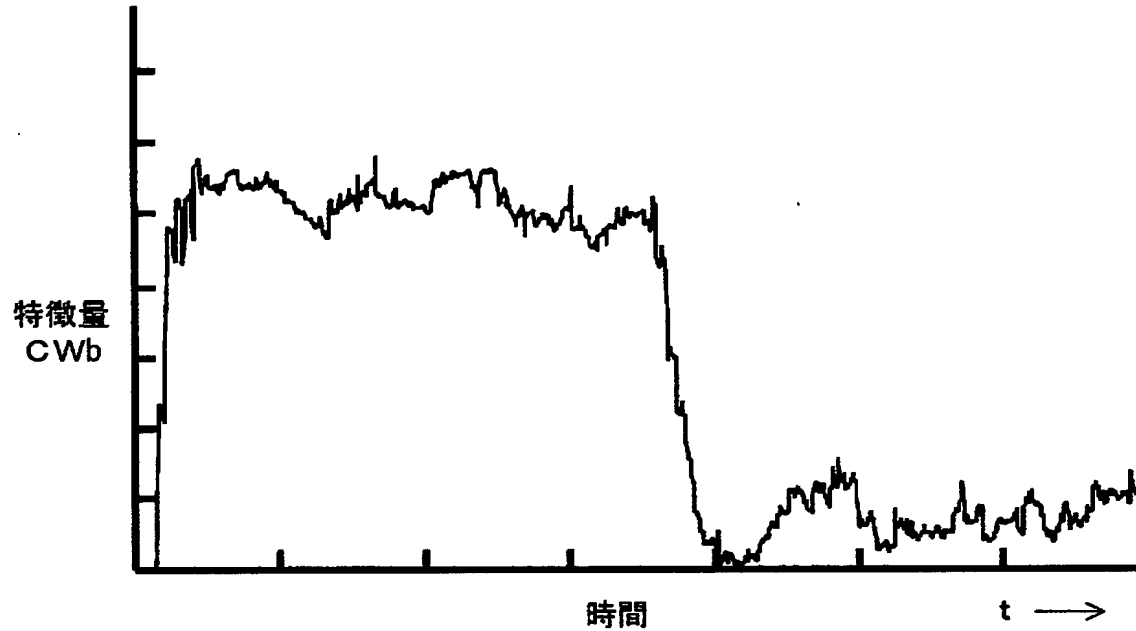
【図 3 5】

# 平均動き量のフーリエ変換結果



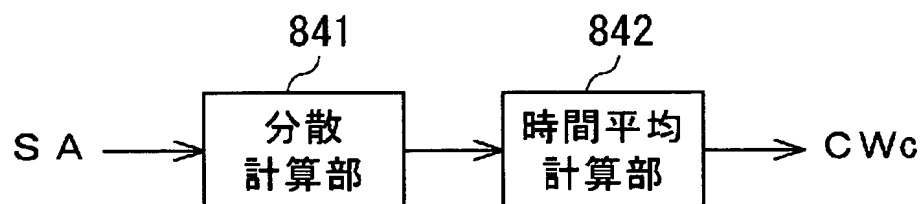
【図 3 6】

### 特徴量 CWb の推移



【図 3 7】

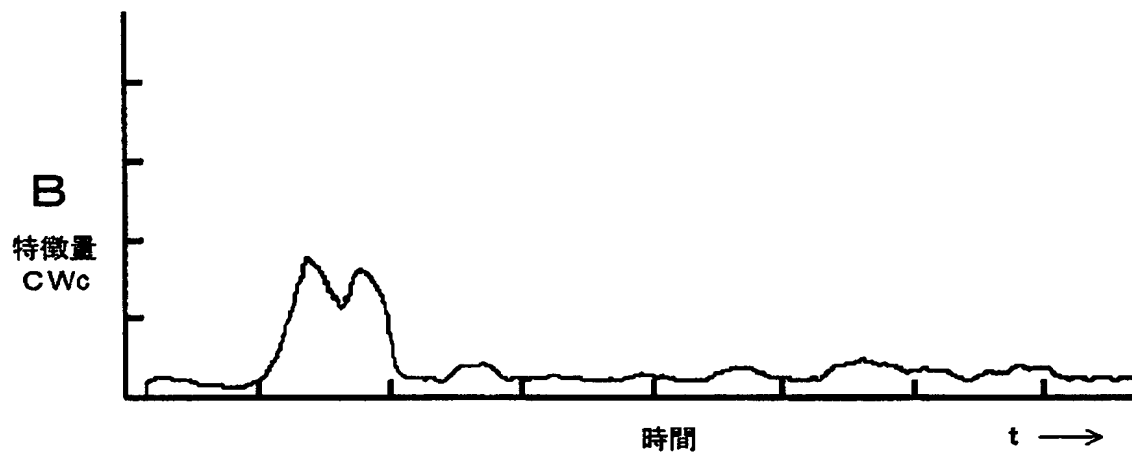
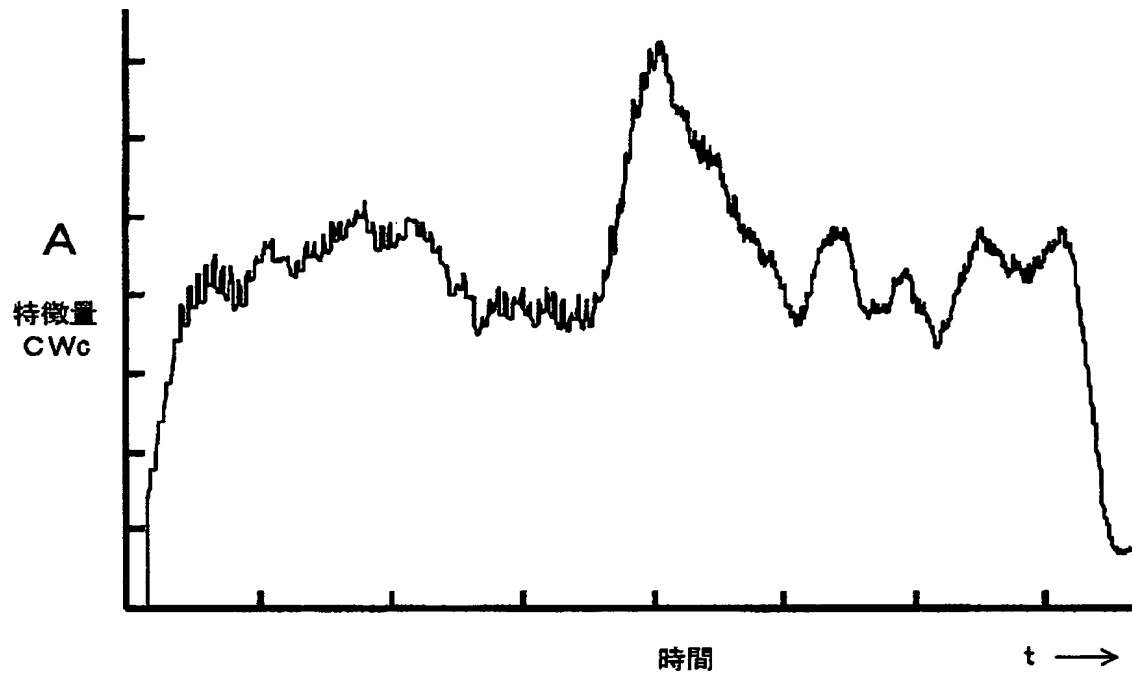
### 音量検出部の構成





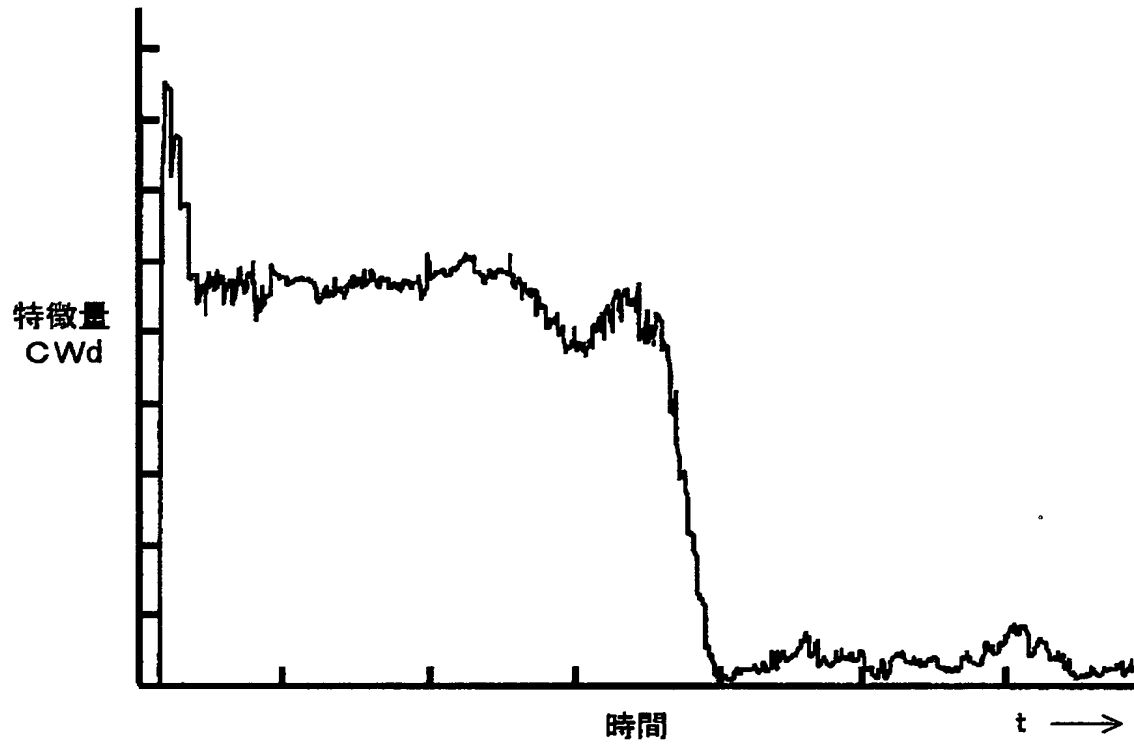
【図 38】

# 特徴量CWcの推移



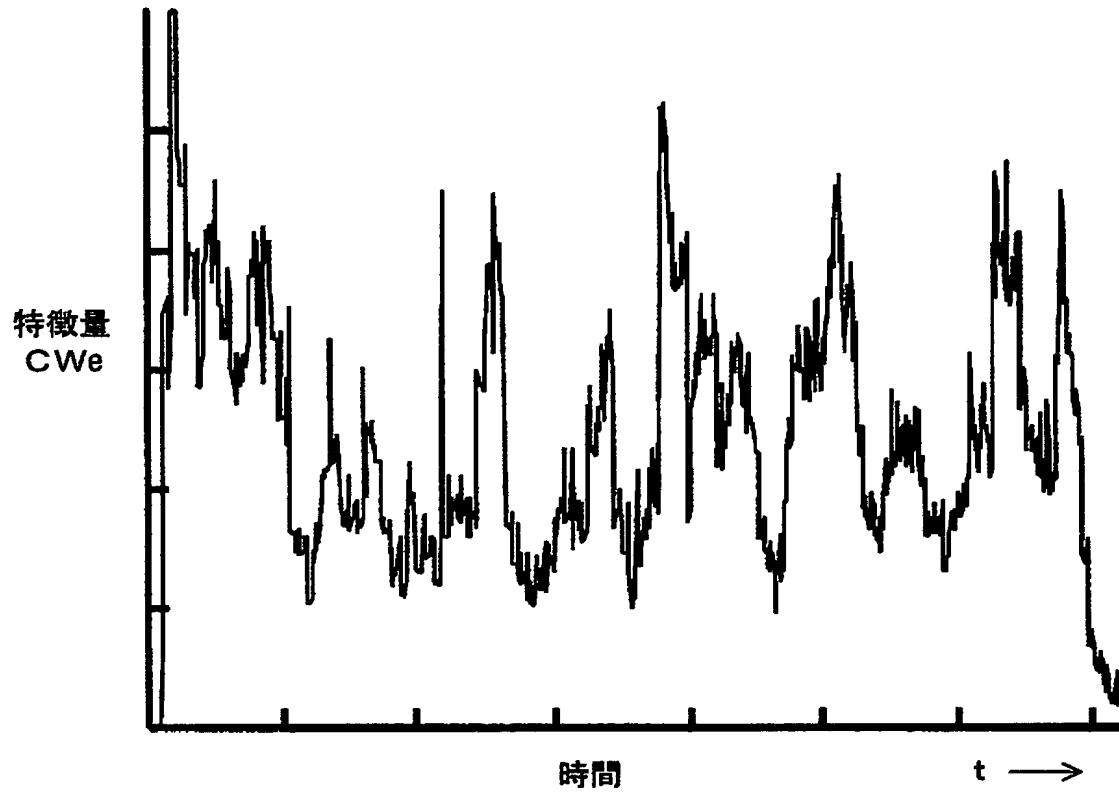
【図 3 9】

### 特徴量 CWd の推移



【図 4 0】

## 特徴量 CWe の推移



【図 4 1】

推定動作 (特徴量  $CW_e$  を用いない場合)

	$CW_a$	$CW_c$
第 1 基準レベル より大	「拍手」あるいは 「手拍子」	
第 1 基準レベルと 第 2 基準レベル の範囲内	「笑い」	
第 2 基準レベル より小	「その他の動作」	

A

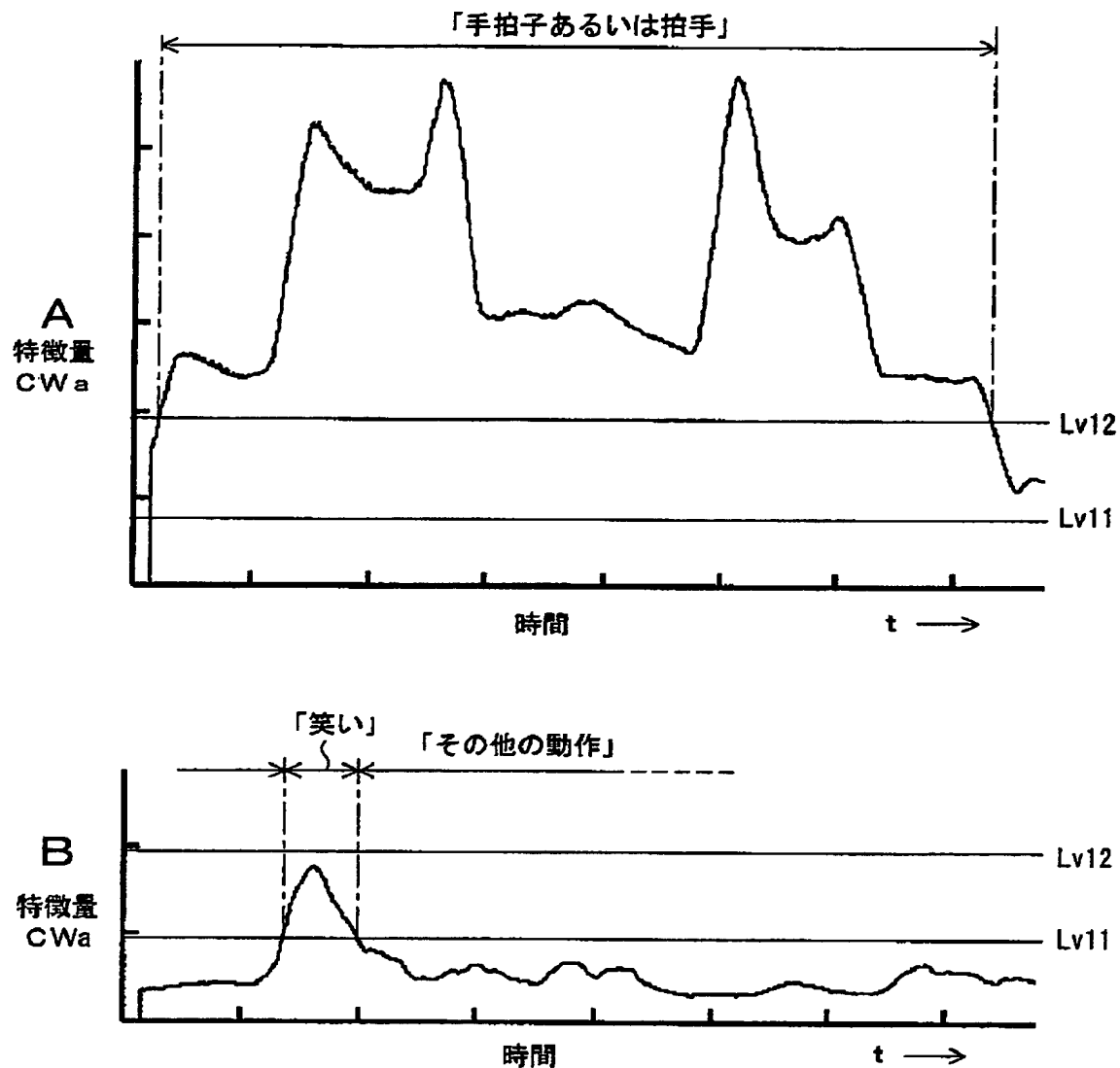


	$CW_b$	$CW_d$
基準レベル よりも大	「手拍子」	
基準レベル よりも小	「拍手」	

B

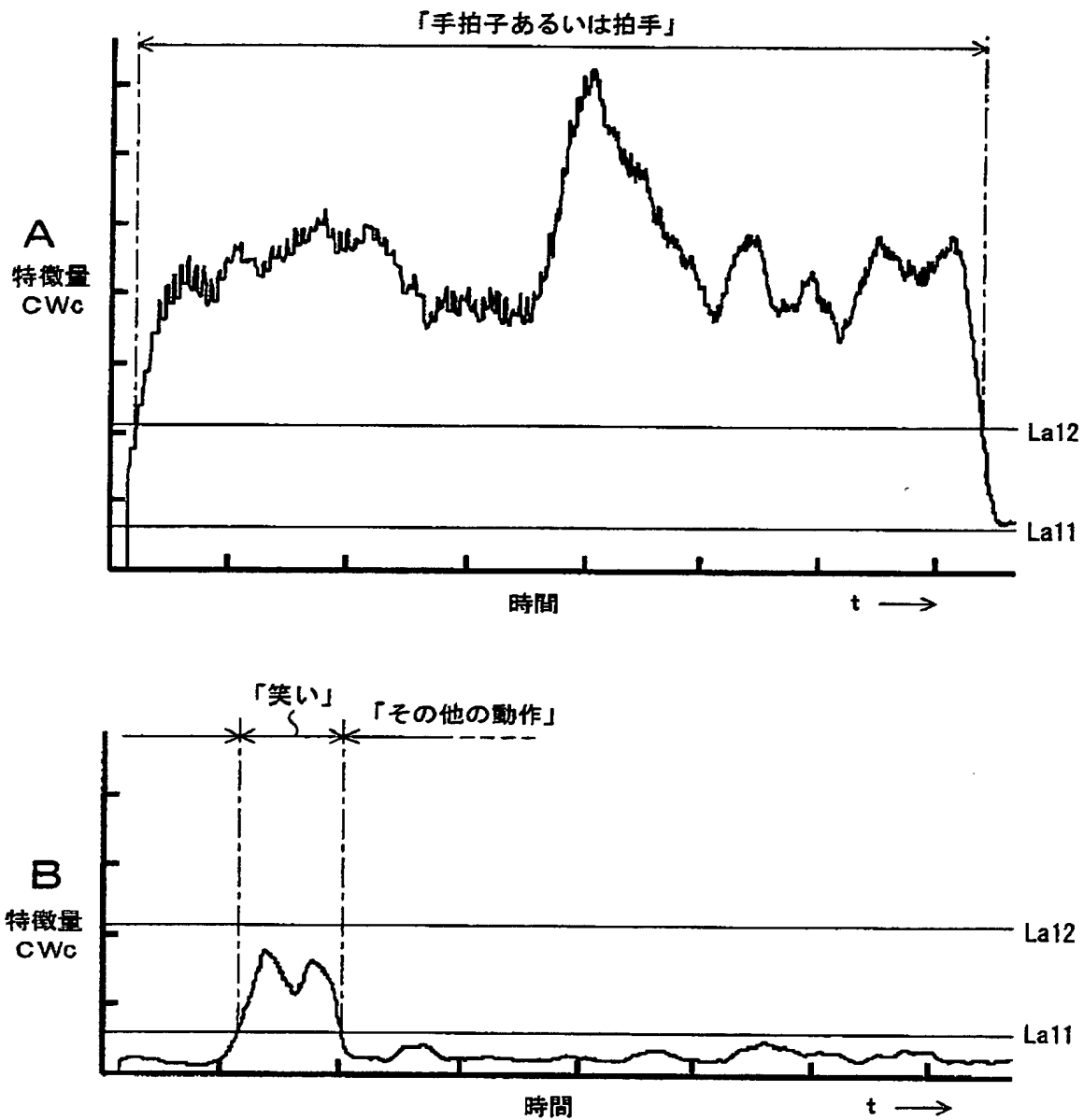
【図 4 2】

# 特徴量に基づいた推定結果（その 3）



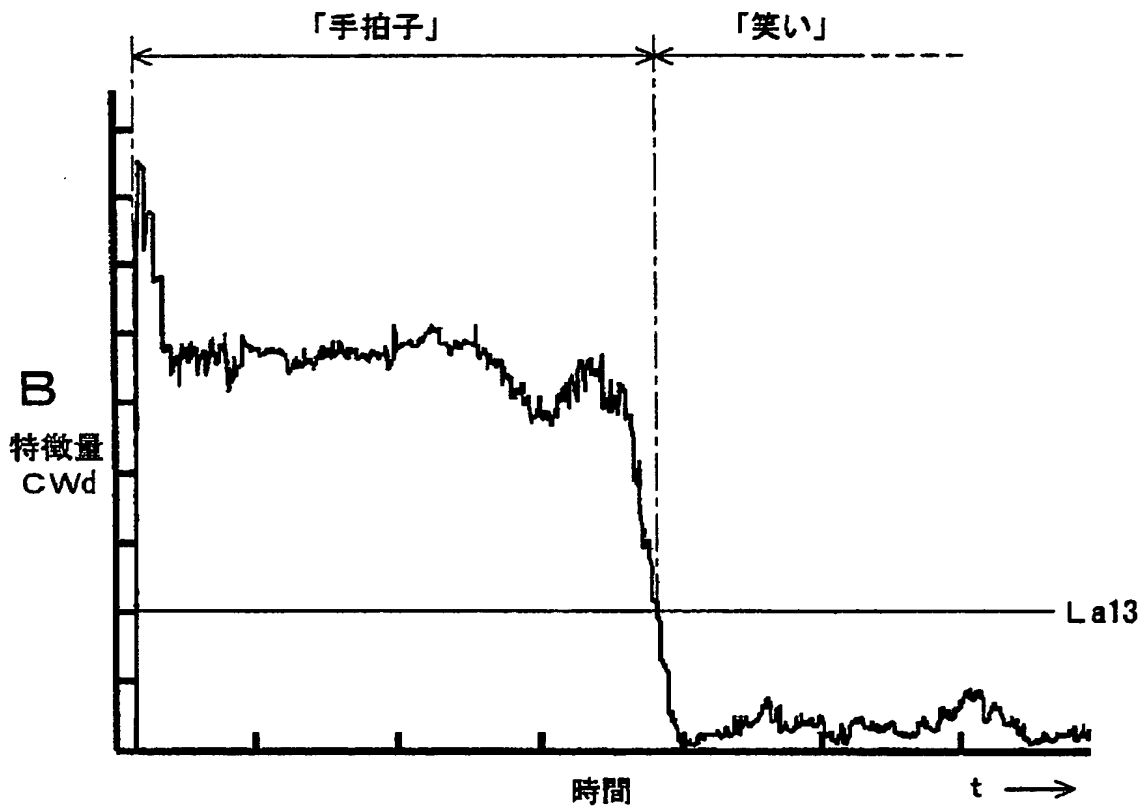
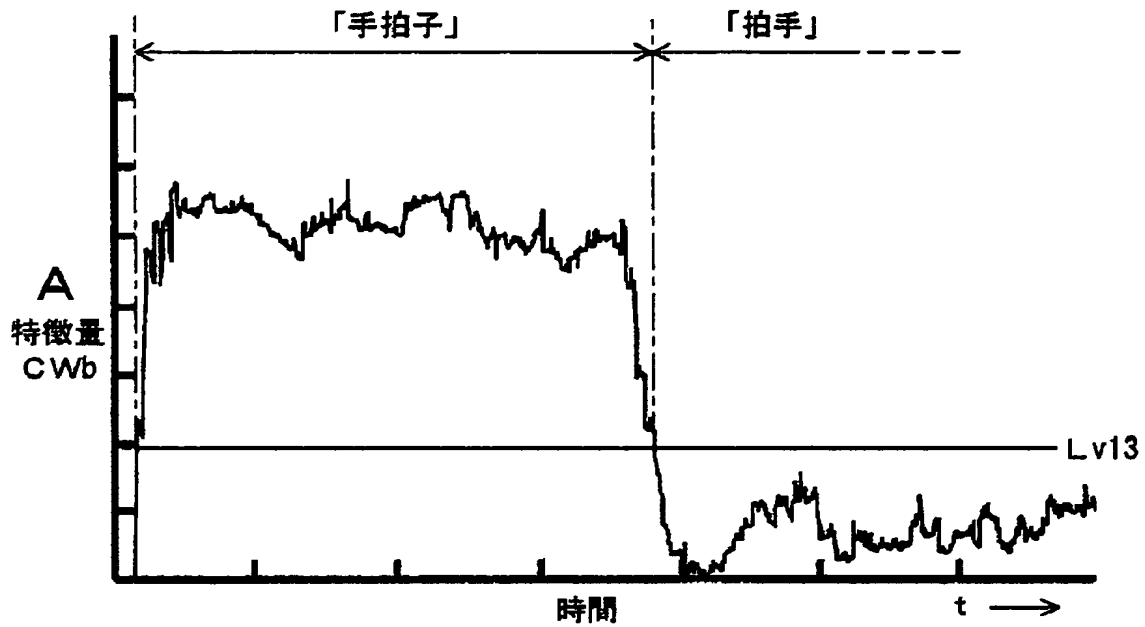
【図 4 3】

特徴量に基づいた推定結果（その 4）

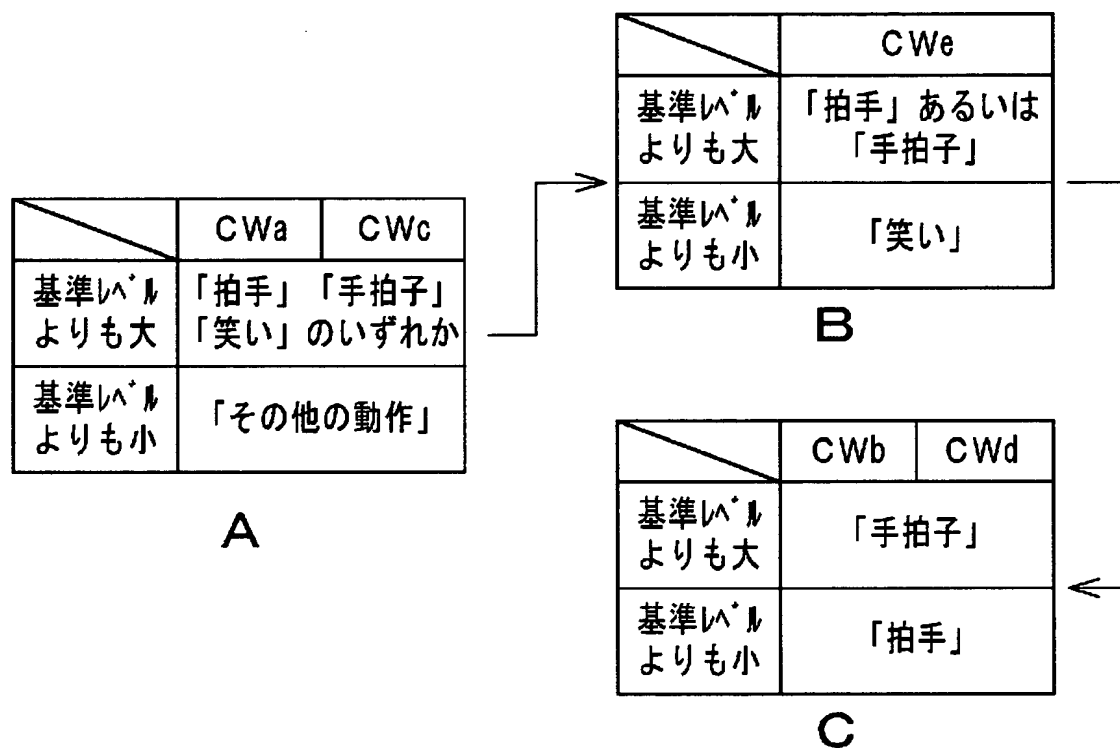


【図 4 4】

特徴量に基づいた推定結果（その 5）

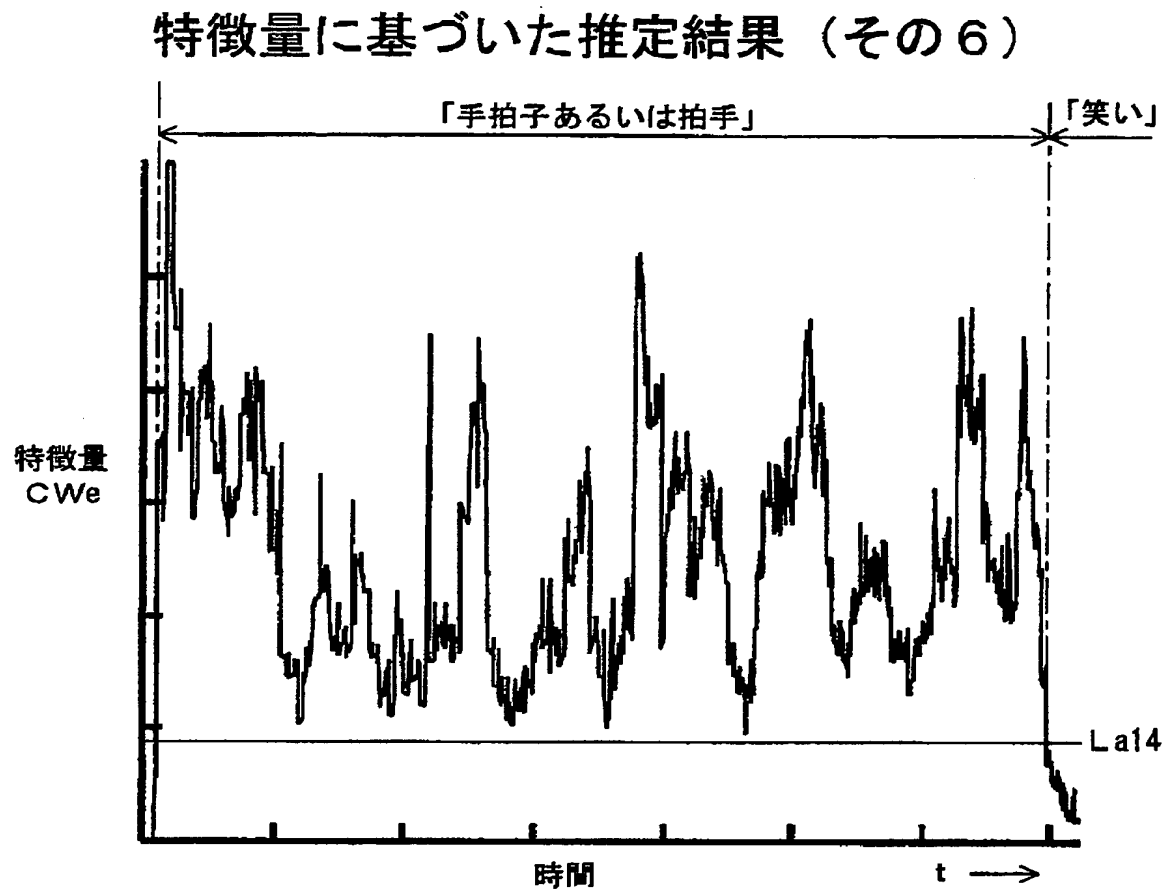


【図 4 5】

推定動作 (特徴量  $CW_e$  を用いる場合)

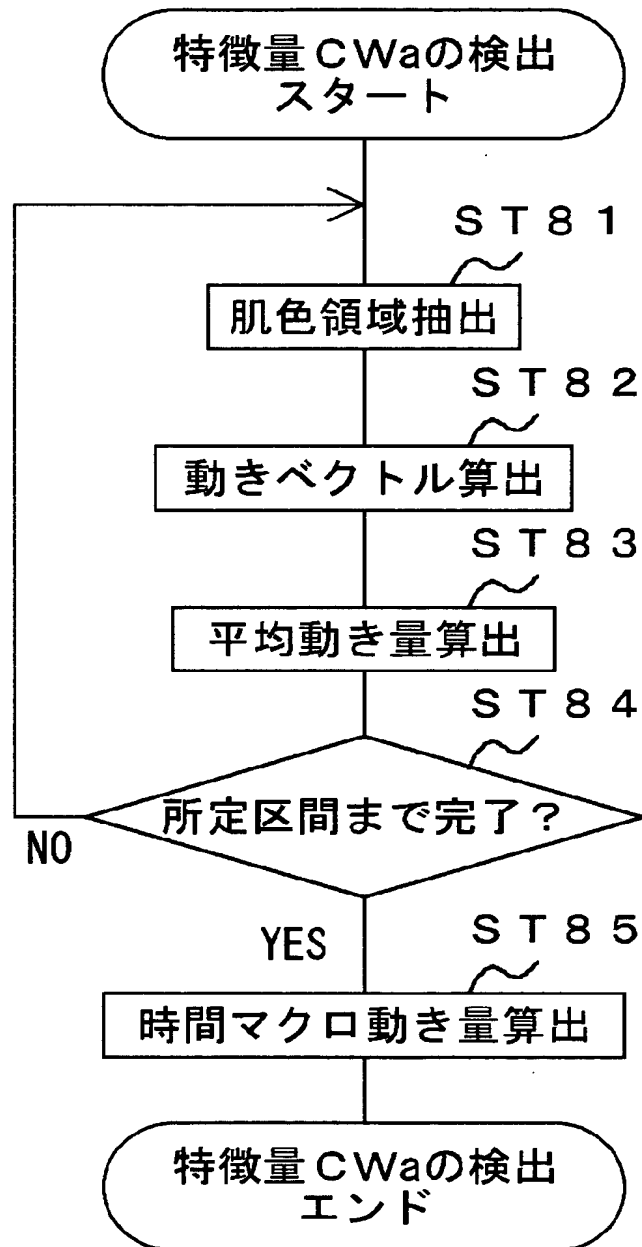


【図 46】



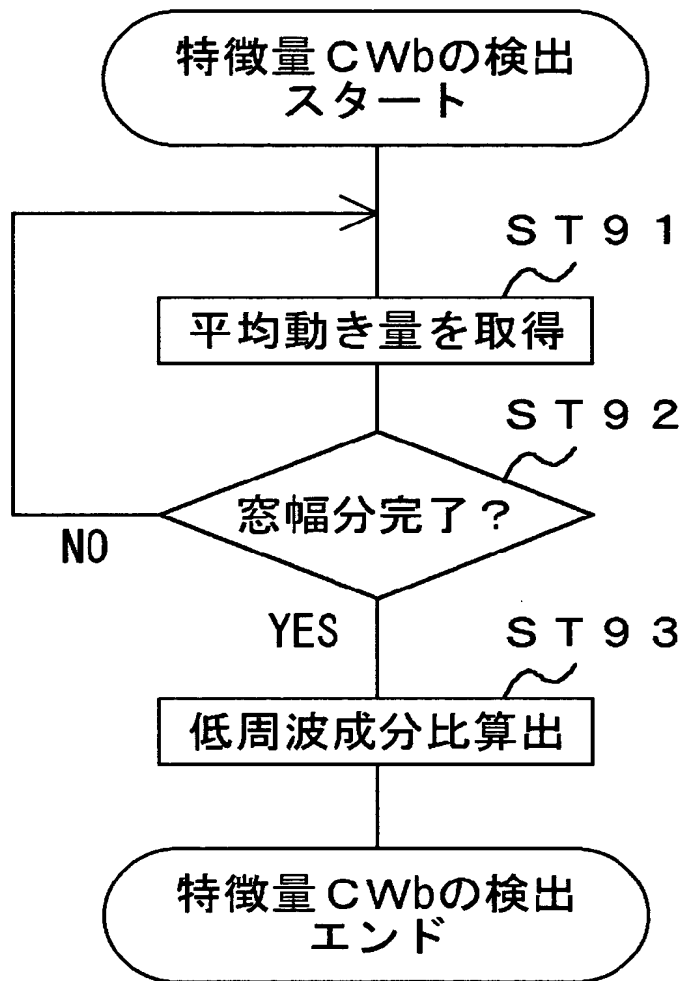
【図 4 7】

## 特徴量 C W a の検出処理



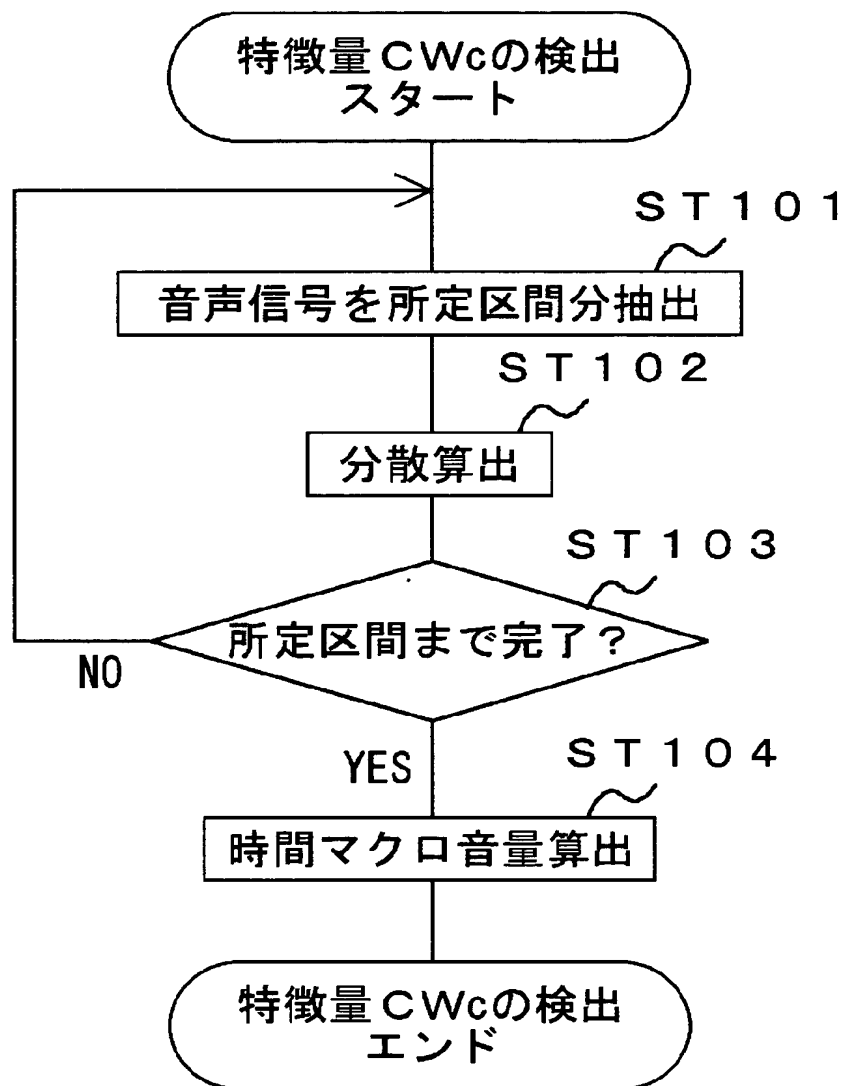
【図 4 8】

## 特徴量 CWb の検出処理



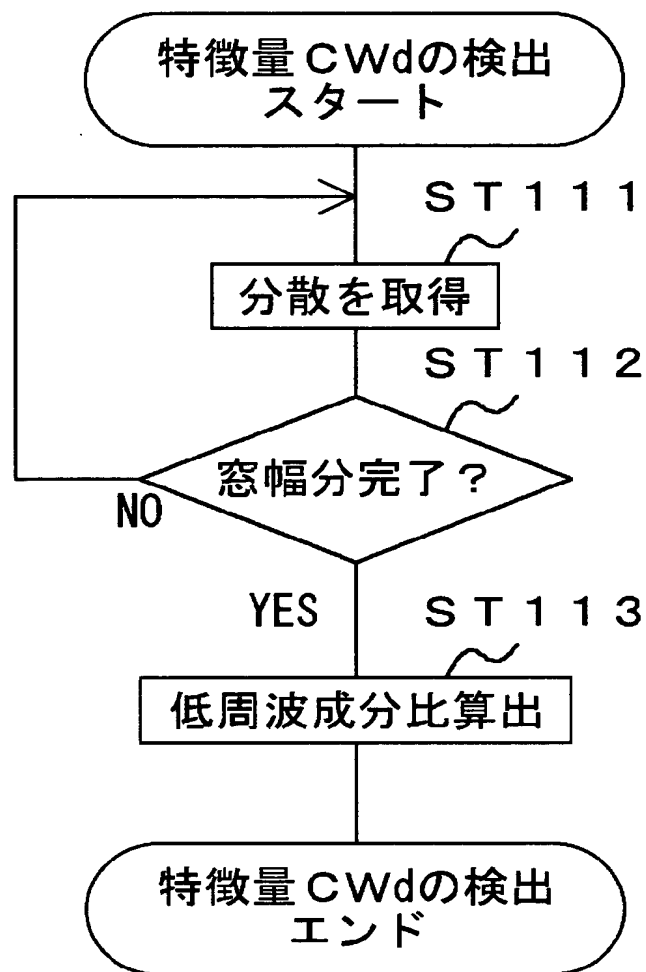
【図 4 9】

## 特徴量 CWc の検出処理



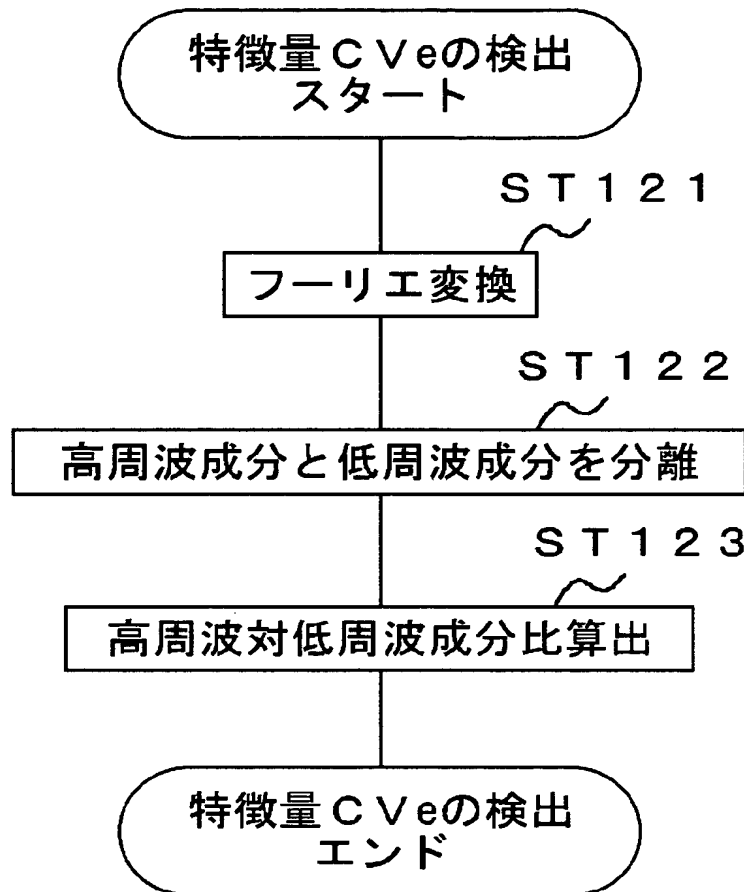
【図 5 0】

## 特徴量 CWd の検出処理



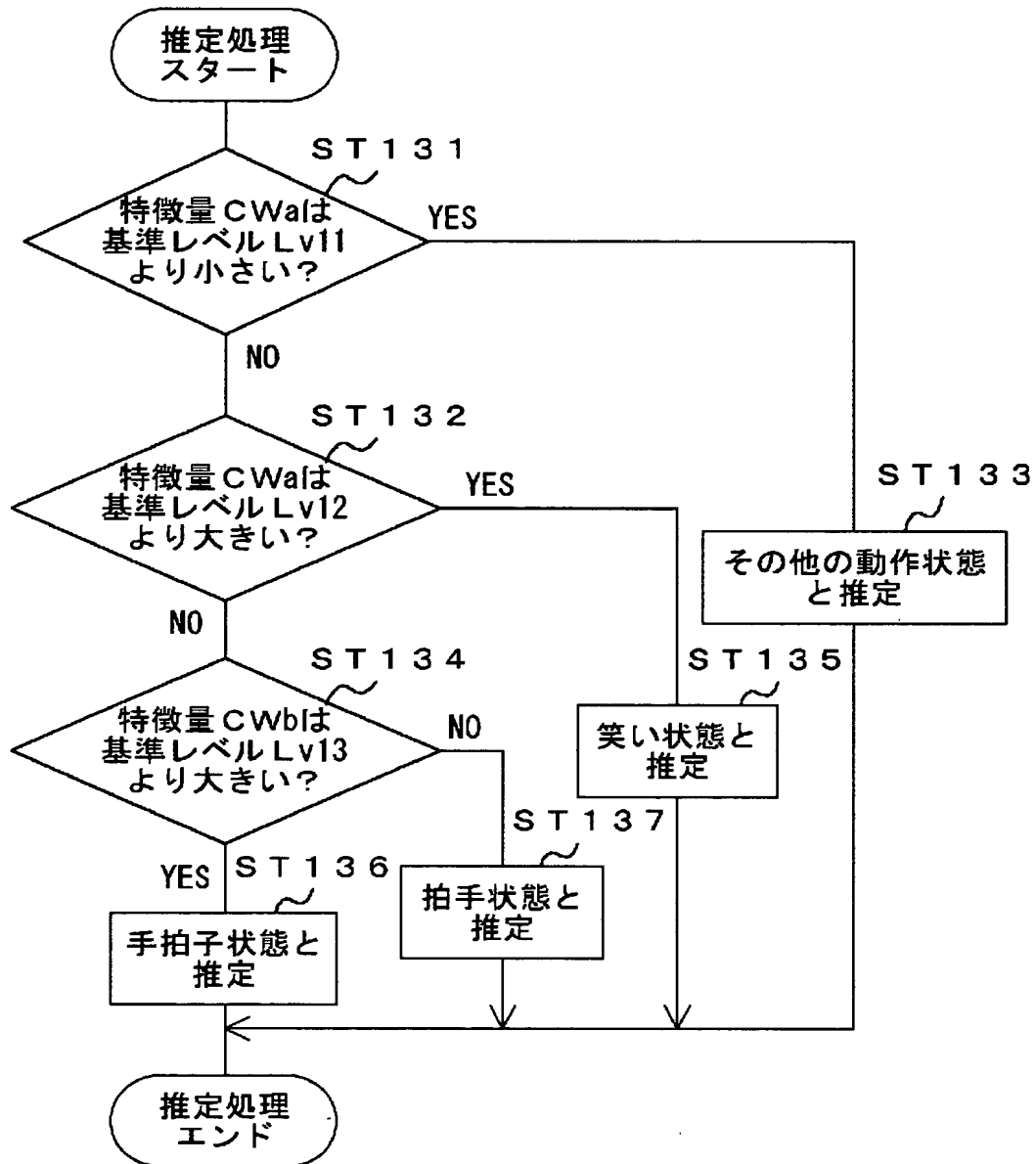
【図 5 1】

## 特徴量 C V e の検出処理



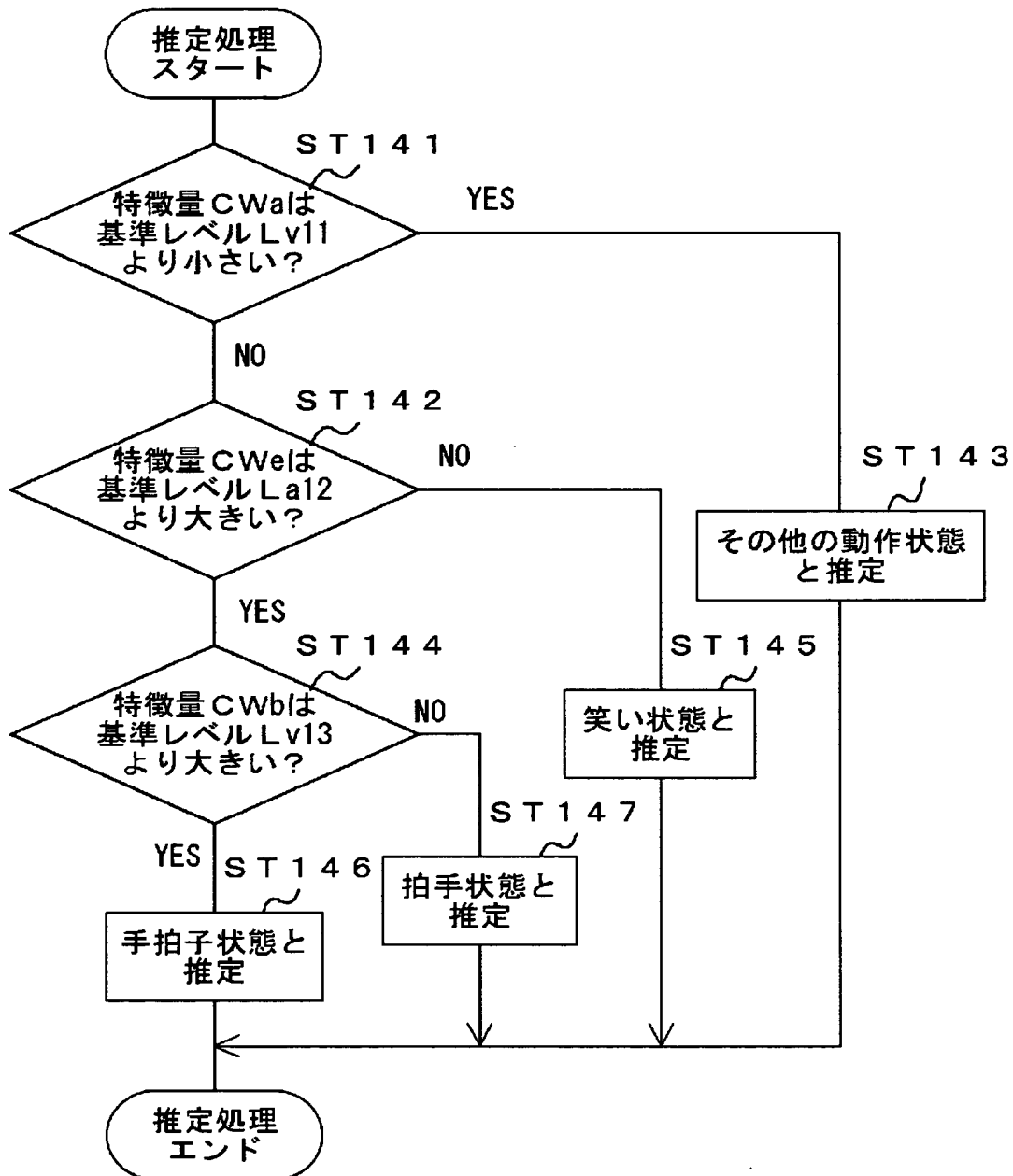
【図 5 2】

# 観客状況の推定処理 (特徴量 CWe を用いない場合)



【図 5 3】

## 反応動作の推定処理 (特徴量 CWe を用いる場合)





【書類名】            要約書

【要約】

【課題】 人が観客状況を観察しなくとも、観客状況を容易に推定できるものとする。

【解決手段】 観客を撮影した画像信号 S V や観客からの音声に応じた音声信号 S A を入力部 2 0 で生成する。特徴量検出部 3 0 ( 8 0 ) では、信号 S V や信号 S A に基づいて、例えば観客状況の動き量や動き周期性、音量、音声の周期性、音声の周波数成分に関する情報を検出する。この検出結果に基づき推定部 4 0 ( 9 0 ) で観客状況を推定する。出力部 5 0 では、観客状況の推定結果を出力する。

【選択図】            図 1

【書類名】 手続補正書

【提出日】 平成14年 7月 2日

【整理番号】 0290049715

【あて先】 特許庁長官殿

【事件の表示】

    【出願番号】 特願2002-186690

【補正をする者】

    【識別番号】 000002185

    【氏名又は名称】 ソニー株式会社

【代理人】

    【識別番号】 100090376

    【弁理士】

    【氏名又は名称】 山口 邦夫

    【電話番号】 03-3291-6251

【プルーフの要否】 要

【手続補正 1】

    【補正対象書類名】 図面

    【補正対象項目名】 図 5 1

    【補正方法】 変更

    【補正の内容】 1

【手続補正 2】

    【補正対象書類名】 図面

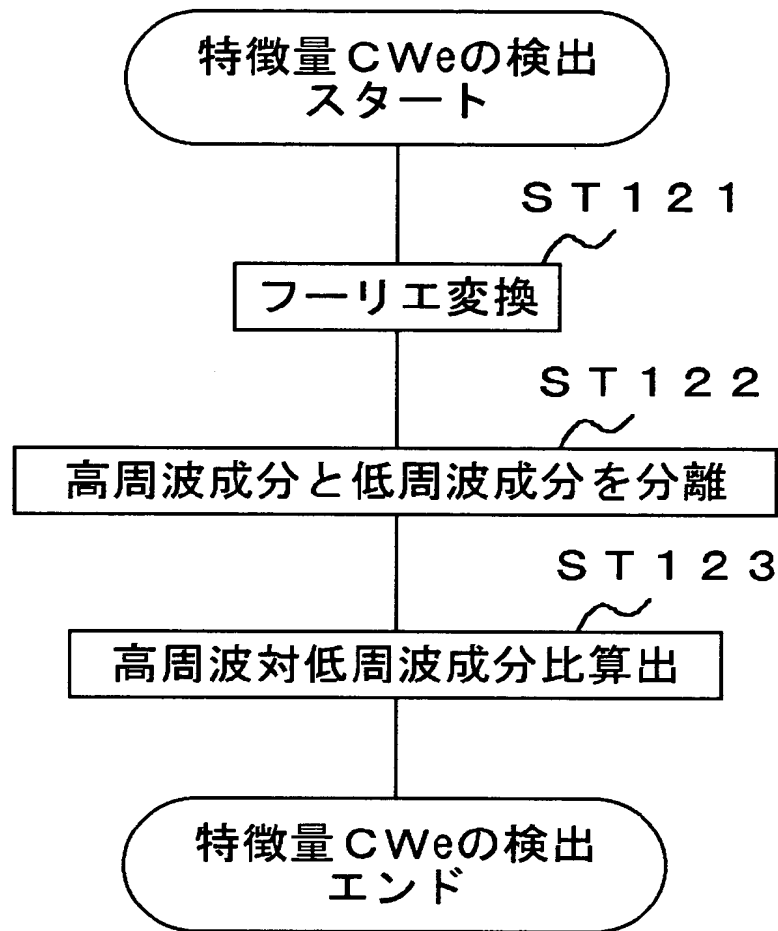
    【補正対象項目名】 図 5 3

    【補正方法】 変更

    【補正の内容】 2

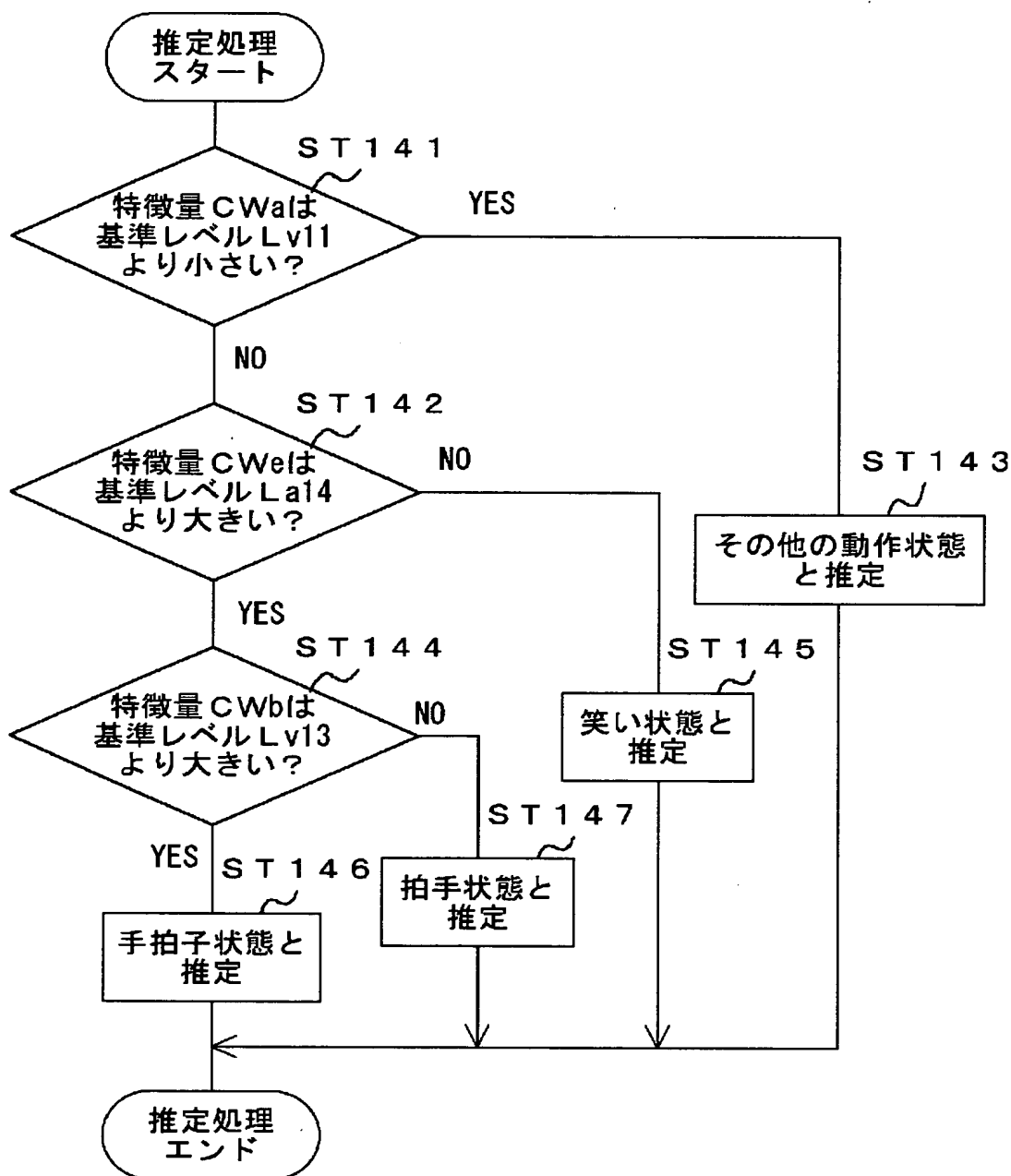
【図 5 1】

## 特徴量 CWe の検出処理



【図 5 3】

# 反応動作の推定処理 (特徴量 CWe を用いる場合)



認定・付加情報

特許出願の番号	特願 2002-186690
受付番号	50200970098
書類名	手続補正書
担当官	塩野 実 2151
作成日	平成14年 7月 8日

<認定情報・付加情報>

【補正をする者】

【識別番号】 000002185

【住所又は居所】 東京都品川区北品川6丁目7番35号

【氏名又は名称】 ソニー株式会社

【代理人】 申請人

【識別番号】 100090376

【住所又は居所】 東京都千代田区内神田1丁目15番2号 平山ビル5階 山口特許事務所

【氏名又は名称】 山口 邦夫

出 願 人 履 歴 情 報

識別番号 [ 0 0 0 0 0 2 1 8 5 ]

1. 変更年月日 1 9 9 0 年 8 月 3 0 日  
[変更理由] 新規登録  
住 所 東京都品川区北品川 6 丁目 7 番 3 5 号  
氏 名 ソニー株式会社